



## Üst düzey düşünme becerilerinin ölçümüne yönelik bağlam ve soru değerlendirme ölçeklerinin geliştirilmesi

### Development of context and question evaluation scales for measuring higher order thinking skills

Yılmaz Orhun GÜRLÜK<sup>1\*</sup> Mediha KORKMAZ<sup>2</sup> Tahsin Oğuz BAŞOKÇU<sup>3</sup> <sup>1</sup> İzmir Demokrasi Üniversitesi, İzmir, Türkiye<sup>2,3</sup> Ege Üniversitesi, İzmir, Türkiye

**Özet:** Eğitimde üst düzey düşünme becerilerin ölçülmesinde ve değerlendirilmesinde modern eğitim anlayışı bağlamlara dayanmaktadır. Bu bağlamlar salt bilginin kullanılmasını değil transfer, analitik çözümleme, eleştirel düşünme gibi çeşitli üst düzey düşünme becerilerin de kullanılmasını gerektirmektedir. Öğrencilerin farklı bağlamlar arasında bilgi transferi yapması, bağlamdaki bilgiler üzerinde stratejiler geliştirerek analitik olarak çözümlemesi ve farklı çözüm alternatifleri geliştirebilmesi beklenmektedir. Üst düzey düşünme becerileri ölçen, açık, anlaşılır ve düzeye uygun bağlamlar kaliteli bağlamlar olarak değerlendirilirken, bu bağlam ile karşılıklı bağımlılık gösteren, bağlama uygun, açık ve anlaşılır sorular da kaliteli olarak değerlendirilmektedir. Bu araştırma "Çevrimiçi Bilişsel Tanıya Dayalı İzleme Modelinin Üst Düzey Düşünme Becerilerine Etkisi" başlıklı 120K850 numaralı TÜBİTAK 1001 projesi kapsamında yürütülmüş ve ilk yazarın doktora tez çalışmasından türetilmiş olup, uzmanların niceliksel olarak bağlam ve sorularını değerlendirebileceği ölçekler geliştirilmiştir. Bağlam kalitesini değerlendirme ölçeğinin Cronbach alfa değeri 0.81 olarak belirlenirken doğrulayıcı faktör analizinde uygun uyum indekslerine ulaşıldığı gözlenmiştir;  $\chi^2/sd = 2.58$ , RMSEA = 0.09, SRMR = 0.05, CFI = .99, GFI = 0.97. soru kalitesini değerlendirme ölçeğinin de Alfa değerinin 0.84 olduğu tespit edilirken uyum indekslerinin kabul edilebilir sınırlarda olduğu görülmüştür;  $\chi^2/sd = 4.49$ , RMSEA = 0.08, SRMR = 0.03, CFI = .99, GFI = 0.96. İki ölçek arasında orta düzeyde pozitif yönlü bir ilişki tespit edilmiştir;  $r = 0.57$ ,  $p < 0.001$ .

**Anahtar Sözcükler:** Üst düzey düşünme becerileri, eğitim, bağlam ve soru kalitesi, ölçek

**Abstract:** In measuring higher-order thinking skills, modern education emphasizes contexts requiring knowledge use and various skills like transfer, analytical solution, and critical thinking. Students must transfer knowledge across contexts, develop strategies, and analyze information analytically, creating different solutions. Quality contexts are clear, understandable, and class-appropriate, with interdependent questions that are also clear and context-appropriate. This research, conducted under the TÜBİTAK 1001 project 120K850, and derived from the first author's PhD thesis, developed scales for experts to evaluate context and questions quantitatively. Principal components analysis confirmed both scales are unidimensional. The Cronbach's alpha value of the context quality assessment scale consisting of 6 items was 0.81, while it was observed that appropriate fit indices were reached in the confirmatory factor analysis;  $\chi^2/sd = 2.58$ , RMSEA = 0.09, SRMR = 0.05, CFI = .99, GFI = 0.97. The reliability value of the question quality assessment scale consisting of a total of 7 items was found to be 0.84, while the fit indices were also within acceptable limits;  $\chi^2/sd = 4.49$ , RMSEA = 0.08, SRMR = 0.03, CFI = .99, GFI = 0.96. A moderate positive correlation was found between the two scales;  $r = 0.57$ ,  $p < 0.001$ .

**Keywords:** Higher order thinking skills, education, context and question quality, scale

## 1. Giriş

Öğrenme becerilerini, öğrenme sürecini ele almada, anlamada ve öğrenilen becerilerin kontrol edilmesinde üst düzey düşünme becerilerin değerlendirilmesi çağdaş eğitim paradigmalarının önemli bir parçasıdır. Bu beceriler ile öğrencilerin yetenek kestirimleri gerçekleştirilmekte ve geleneksel olarak bilginin doğrudan kullanılması değil dolaylı olarak

sentezlenebilmesi ve sentezlenene bu bilgiler aracılığıyla yeni bilgilerin üretilebilmesi ana amacı oluşturmaktadır (Biggs & Tang, 2011; Bransford & Schwartz, 2001; Brandford, Brown & Cocking, 2000); böylece öğrencilerin saf olarak bilgiyi kullanma düzeyleri değil ayrıca yeni çözümler bulabilme, stratejiler geliştirme becerileri ölçülmekte (Wang ve Chen, 2023; Yu, Wang ve Yuizano, 2020), ek olarak bilişsel kapasiteleri hakkında yorum yapma imkânı doğmaktadır (Alami vd., 2020; Elballah vd., 2024; Brandford, Brown & Cocking, 2000). Öğrencilerin bilgileri bağlamlar ve durumlar arasında transfer etmesi, soruların veya bağlamların analitik stratejiler kullanılarak çözümlenmesi ve bilgi parçalarına eleştirel bir bakış sunulabilmesi gibi birtakım yeterlikler üst düzey düşünme becerilerine ait ölçümlere odaklanan sorularda karşımıza çıkmaktadır. Bu becerileri ölçmek ve geliştirmek amacıyla eğitimde kullanılan ölçme araçlarının ilerlemesi için öğrencilerin saf bilgi ile çözümü bulması değil gerçek dünya durumlarını daha iyi yansıtan karmaşık bilgilerle çözüme ulaşması gerekmektedir. Bu becerilerden ilki olan transfer becerisi, bir öğrencinin bir bağlam ya da konuda öğrendiği bilgi ve becerileri farklı durum, konu ya da bağlamlara uygulama becerisini ifade etmektedir. Transfer becerisi, öğrencinin, öğrenmeyi genelleştirebilmesini ve yeni bir soru ya da sorunla karşılaştığında geçmişteki çözüm yollarını yeni duruma aktarabilmesini göstermektedir. Farklı durumlar içerisinde bilginin kullanılabilmesi bilginin ezberlenmeden, içselleştirildiği anlamına gelmektedir. Diğer taraftan transfer becerisi gelişmiş bir bireyin yaşam boyu öğrenme perspektifinin daha yüksek olması beklenmektedir (Barnett & Ceci, 2002; Butterfield, Nelson & Peck, 1988). Bilgileri başka bağlamlara transfer edebilen bireyler, yeni sorunlar karşısında daha hızlı ve yaratıcı çözümler sunabilmektedir. Transfer becerisine sahip olanlar içerikleri farklı bağlamlar arasında aktarabilir, başka bir alanda öğrendiği bilgiyi yeni bir alanda kullanabilir ya da bilgileri başka formatlarda (örneğin: metinden tabloya dönüştürerek) ifade edebilir. Bunun yanı sıra başka bir süreç içerisinde edindiği bilgiyi günlük hayatta ya da eğitimde başka bir sürece dahil edebilir (Barnett & Ceci, 2002). Transfer becerisi farklı bilgi parçalarının iç içe geçirerek yeni ve kompleks bilgiler üretmenin de bir parçasıdır. Bağlamlarda yer alan karmaşık ilişkiler transfer becerisi ile ortaya çıkartılarak örtük yapılar veya farklı bilgi parçaları anlaşılabilir. Özetle transfer becerisi, bilginin doğrudan başka bağlama aktarılmasını; basit bilgiden aktif çözüm sürecinde karmaşık bilgi ortaya çıkartılmasını ve kapsamlı bir stratejinin farklı bir süreçlerde kullanılmasını kapsar (Halpern, 2013; Jonassen, 2000). Transfer becerisinin bilinçli olarak kullanılması akademik performans arttırmakta ve artan akademik performans da beceriyi günlük yaşamda kullanmaya teşvik etmektedir (Bransford & Schwartz, 2001).

Analitik çözümleme becerisine gelindiğinde ise öğrencilerin soruları daha kolay çözümlenebilir alt parçalara ayırma yeteneği, sonuçlara sağlam gerekçelerle ulaşma ve çözüme dair kanıtları sistematik olarak değerlendirebilme yetileri değerlendirilmektedir. Özellikle matematik ve fen eğitimi gibi daha karmaşık çözümler gerektiren eğitim alanlarında ön plana çıkmaktadır (Pintrich, 2002; Tobias & Everson, 2000). Mühendislik gibi teknik ve mantıksal yaklaşımlar gerektiren disiplinlerin ayrılmaz bir parçası olan analitik çözümleme becerisinin belirlenmesi öğrencilerin akademik ilerlemelerinde uygun yönelimi bulmaları bakımından da önemlidir (Tobias & Everson, 2000, Sternberg, 1986). Öğrencileri analitik olarak çözümleri bulmaya yönlendirmek, bu becerilerin kaçınılmaz olduğu profesyonel alanlara hazırlanmayı beraberinde getirmektedir. Analitik çözümleme becerisi de yalnızca akademik bir beceri olmayıp günlük hayatta da işlevsel olarak kullanılan bir beceridir. Gerçek dünyada karşılaşılan sorunların uygun parçalara bölünüp ele alınabilmesi sorunun çözümünü kolaylaştırmaktadır (Halpern, 2013; Jonassen, 2000). Analitik çözümleme becerisine sahip bir birey çalışmanın ya da soruların kurallarını belirleyerek mevcut ilişki ve işlemleri tanımlayabilir. Sorunların ele alınmasında yeni yaklaşımlar sunabilir ve eylemlerin/çözümlerin potansiyel sonuçlarını önceden tespit edebilir. Analitik düşünme becerisi bilgi kümelerinin tanımlanmasını, çözüme yönelik hızlı ve uygun bilgiler toplanmasını, temel unsurların tespit edilmesini kapsar. Olaylar ve süreçlerin arasındaki neden sonuç ilişkisini daha kolay kavrayan birey, olası seçenekleri daha iyi

değerlendirerek yeni çözümler belirler, bir duruma veya soruna ait çoklu nedensellikleri belirleyebilir, risk tahminlerinde bulunarak eylem planları hazırlar ve yeni çözüm yolları üretir (Facione, 1990; Halpern, 2013). Analitik düşünme becerisinin günlük yaşamın da temel bir becerisi olduğu düşünüldüğünde öğrencilerin geleceğe hazırlanabilmesi için ele alınması gerektiği bilinmektedir; analitik becerileri yüksek olan öğrencilerin iş bulma performansları, yeni okullarda kendini gerçekleştirme becerileri ve bilişsel düzeyleri diğer öğrencilerden ayırmakta ve böylece farklı perspektifler geliştirebilmektedirler (Baysal ve Ocak, 2022; Beno ve Koacova, 2020; Hidayat vd., 2024; Krell, Kind ve Krüger, 2024)

Eleştirel düşünme becerisi ise bu temel üç becerinin hiyerarşik olarak en üstünde yer almaktadır. Var olan argümanları değerlendirme, argümandaki çelişkili tarafları tespit edebilme, önyargıları yaratan parçaları belirleme, muhakeme ile çözümü bulma ve kanıtlara dayandırarak ilerleme gibi aşamaları içermektedir. Bu üst düzey düşünme becerileri, bilgilerin güvenilirliğini değerlendirmeyi de kapsamına alarak yanlış veya hatalı bilgileri ayırt etmeyi ve yargılamayı kolaylaştırmaktadır. Eleştirel düşünme becerisi aynı zamanda bağımsız düşünme becerisi olarak da ele alınmaktadır. Bireyin kendi başına bilgileri sorgulayabilme düzeyini ifade etmektedir. Güçlü eleştirel düşünme becerilerine sahip bireyler karmaşık sorunları ele almakta ve bilinçli kararlar vermekte diğerlerinden daha iyi performans göstermektedir (Veenman, Wolters & Afflerbach, 2006). Eleştirel düşünme becerisi beraberinde bilgiye dair boşlukları tespit etme becerisi de getirir. Böylece bireyler bir konu hakkında yargıda bulunmadan önce o konu hakkındaki varsayım ve motivasyonları tanımlayabilir. Metin veya konuşma içerisindeki akıl yürütmelerin belirlenmesi, gerekçelendirme sürecindeki kusurların saptanması, yanlışlıkların ortaya çıkartılması gibi birçok beceriyi içeren alan eleştirel düşünme becerisi kişinin kendi fikirlerini, argümanlarını ya da çözüm yollarını belirleyebilmesini sağlar. Argümanları açığa çıkarmak ve güçlendirmek için karşılaştırmalar yaparak kararlar alır ve kararlara ilişkin olası sonuçları önceden kestirebilir (Paris & Winograd, 1990; Tobias & Everson, 2000).

Üst düzey düşünme becerilerinin eğitim programlarına dahil edilmesi sadece saf bilginin değil aynı zamanda düşünmenin parçası olan yeteneklerin de ele alınmasını sağlamaktadır. Çağdaş sistemler ve programlar daha fazla bilişsel performans gerektiren bağlamlara dayalı soruları, doğrudan bilginin ölçüldüğü sorulara göre daha fazla tercih etmektedir (Chi ve diğ., 1989; Gulikers, Bastiaens & Kirschner, 2004). Böylece sorularda daha fazla üst düzey düşünme becerileri ölçülmektedir. Bilgiyi doğrudan kullanan değil üst düzey düşünme becerileri ile bilgi parçalarını sentezleyerek kullanan öğrencilerin akademik ve günlük yaşam performanslarının daha iyi olduğu bilinmektedir. Üst düzey düşünme becerileri öz-düzenleme yeteneğini de arttırmaktadır. Dolayısıyla öğrencilerin üst düzey düşünme becerilerinin geliştirilmesine yönelik sistemler günlük yaşamı da etkilemektedir (Schraw & Moshman, 1995). Üst düzey düşünme becerilerinin geliştirilmesi ve değerlendirilmesi için farklı sınıf ve yetenek düzeylerinde ölçüm araçlarına ihtiyaç artmakta ve buna yönelik çalışmalar hız kazanmaktadır. Ancak bağlamların ve soruların bu becerileri ne kadar değerlendirebildiğinin değerlendirilmesi üzerine literatür oldukça sınırlıdır. Öte yandan; bağlamların ve bağlama bağlı soruların kalitesi üst düzey düşünme becerilerinin uygun bir şekilde ölçülmesini belirlemektedir (Kuhn, 1999; Paul & Elder, 2008). Bağlam, öğrenmeye ait görevlerin yerleştirildiği temel parçayı ifade ederken, buna bağlı sorular, öğrencilerin bu çerçeve içindeki bilgileri ne kadar kullanabildiği, sentezleyebildiği, yeniden üretebildiği gibi becerilerin ölçülmesine odaklanır. Sorular bağlamlardan, bağlamlar da sorulardan bağımsız olarak ele alınamayacağından transfer, analitik çözümleme ve eleştirel düşünme becerilerinin ölçülmesinde her ikisi de kritik öneme sahiptir (Ennis, 1996; Flavell, 1979; Zimmerman, 2002).

Eğitimde kaliteli bağlamlar, öğrenme süreçlerinin ölçümünde temeli oluşturur. İyi oluşturulmuş bir bağlam, durumsal gerçekliğe uygun olmalıdır. Birçok araştırma, öğrencilerin ilgi çekici, anlaşılabilir ve hayata uygulanabilir materyalleri daha rahat algıladıklarını ve öğrenme düzeylerinin daha etkili olduğunu, daha etkili öğrenme öğrencilerin de becerilerini

günlük hayatta daha iyi sergilediğini ortaya koymaktadır (Bigg & Tang, 2011; Hasanah ve Retnawati, 2022; Rutumalessy vd., 2023; Sadler, 1989, Wood, Bruner & Ross, 1976). Bağlaman kalitesi öğrencilerin sınıf düzeyine uygunluğu, öğrenilen konularla ilgililiği ve önceden bildiklerinin uygulanabilirliği ile ilişkilidir. Ayrıca kaliteli bir bağlam özgün bir şekilde senaryoları ve problemleri tanımlar (Lave & Wenger; 1991; Hattie & Timperley, 2007). Bunların yanı sıra bağlam, öğrencileri becerilerini en iyi şekilde kullanmaya teşvik etmeli; ancak anlamayı güçleştirecek seviyede zor olmamalıdır (Lave & Wenger, 1991; Mayer, 2002).

Bağlama bağlı soruların kalitesi de ölçümün uygunluğunun bir parçasıdır. Bağlamı kullanmada, bilgi ve becerilerini sergilemede ne kadar iyi olacakları sorular üzerinden tespit edilir. Sorular, öğrencilerin bağlamdan ve öğrendiklerinden bilgiyi transfer ettikleri, analitik stratejiler geliştirdikleri ve eleştirel olarak yeni çözümler belirledikleri ana kısımlardır. Yüksek kaliteli bağlama bağlı soruların kalitesi düşük olduğunda bağlaman kalitesi, amaca uygun kullanılamamaktadır. Soruların kalitesini belirleyen en önemli kısımlardan birisi öğrenme hedeflerine uyumlu olmasıdır (Bloom, 1956; Brown, Collins & Duguid, 1989; Bransford, Brown & Coking, 2000). Sorular amaçlanan bilgileri ölçmelidir. İkinci olarak sorular öğrencilerin becerilerini kullanmalarını gerektirerek basit bilgilerin hatırlanmasının ötesine geçmelidir (Mayer, 2002; Perkins & Salomon, 1988).

Bağlamlar gerçek dünya ile uyumlu olarak tasarlanmalı ve eğitim sistemi içerisinde takip edilen programa uygun olmalıdır. Sorular ise açık ve kesin ifadelerle oluşturulmalı ve belirsizlik önlenerek soruların anlanması sağlanmalıdır. Öte yandan bağlama bağlı sorular tek tipte değil farklı türlerde (çoktan seçmeli, çoklu seçmeli, boşluk doldurma, doğru yanlış vb.) olmalıdır (Sadler, 1989; Spiro ve diğ., 1988).

Bu çalışmada uzmanların bağlam ve soruları niceliksel olarak değerlendirebilmesi amacıyla; yukarıda bahsedilen beceriler, bağlamların sınıf düzeyine uygunluğu, bağlamların anlaşılabilirliği gibi noktalara odaklanarak Bağlam Kalitesi Değerlendirme Ölçeği (BKDÖ) ve bağlama bağlı soruların, anlaşılabilirliğinin, sınıf düzeyine uygunluğunun, kalitesinin ele alındığı Soru Kalitesi Değerlendirme Ölçeği (SKDÖ) geliştirilmiştir. Ölçekler, bağlamlarda temsil edilen örtük yapıların ilgili konuya uygunluğunun uzmanlar tarafından belirlenmesinde uzman görüşüne dayalı geçerlik çalışmalarına niceliksel bir yaklaşım olarak oluşturulmuştur. İlgili ölçekler bireylerin ölçümü amacıyla değil; eğitimde ve psikolojide üst düzey düşünmeyi ölçmek için tasarlanmış bağlamlar ve bu bağlamlara dayalı soruların uzmanlar tarafından niceliksel olarak değerlendirilmesi amacıyla oluşturulmuştur. Diğer bir ifade ile bağlam ve sorularının ölçmesi gereken şeyleri ne derecede ölçtüğü saptanmakta; bağlam ve sorularının hem bir arada hem de ayrı ayrı ölçüm koşullarına uygunluğu ele alınmaktadır.

## 2. Yöntem

### 2.1. Bağlam ve Soru Değerlendirme Ölçeklerinin Geliştirilmesi

Ölçeklere ait madde havuzu ilgili projenin paydaşı olan uzmanlar (öğretmenler) tarafından oluşturulmuştur. Çalışmada bu doğrultuda uzmanlar tarafından değerlendirilecek ve eğitim bağlamında üst düzey düşünme becerilerini değerlendirmek için tasarlanmış bağlamlar ile bu bağlamlara dayalı soruların bahsedilen becerileri ölçme, sınıf düzeyine uygun olma gibi çerçevelerden ele alındığı bağlam ve soru kalitesi değerlendirme ölçeklerinin uygulaması Ege Üniversitesi Sosyal ve Beşerî Bilimler Araştırma ve Yayın Etiği Kurulu'nun (Protokol No: 721) 02.12.2020 tarihli 15/01 toplantısında alınan onay kararı ile yürütülmüştür. Hem altı maddeden oluşan bağlam değerlendirme ölçeği hem de yedi maddeden oluşan soru değerlendirme ölçeği için uzmanlardan 5 dereceli Likert ölçeği (1 = hiç uygun değil; 5 = tamamen uygun) ile puanlama yapmaları istenen araştırma, temel araştırma özelliği taşımaktadır.

## 2.2. Örneklem

Bu çalışma için baş yazarın tez çalışması doğrultusunda açımlayıcı faktör analizi (AFA) ve doğrulayıcı faktör analizi (DFA) için ayrı ayrı 581 katılımcıdan oluşan uzmanlar gönüllü olarak bağlam değerlendirmeyi kabul eden kişiler arasından rastgele örnekleme yoluyla belirlenmiştir. AFA örneklemindeki uzmanların görev süresi ortalaması 15.23 (S = 7.23) olarak tespit edilirken; örneklemin %48.88'i kadınlardan (n = 284) ve %51,12'isi erkeklerden (n = 297) oluşmaktadır. Türkiye genelinde örneklemin %49.74'ü (n = 289) Marmara, %23.58'i (n = 137) Ege, %11.02'si (n = 64) İç Anadolu, %7,06'sı (n = 41) Akdeniz, %4.47'si (n = 26) Karadeniz, %3.44'ü (n = 20) Doğu Anadolu ve %0.69'u (n = 4) Güneydoğu Anadolu Bölgelerinden olduğu saptanmıştır. Öğretmenlerin tamamı ortaokul düzeyinde olup 199 öğretmen (%34.25) Türkçe, 187 öğretmen (%32.19) Fen Bilgisi ve 195 öğretmen (%33.563) de Matematik alanında uzmandır.

DFA örneklemini incelendiğinde ise örneklemin görev süresi ortalamasının 13.22 (S = 5.38) olduğu görülmüştür. DFA örnekleminin %51,64'ü kadınlardan (n = 300), %48.36'sı ise erkektir. (n = 281). DFA örneklemini incelendiğinde katılımcıların %49.05'inin (n = 285) Marmara, %22,38'inin (n = 130) Ege, %9.81'inin (n = 57) İç Anadolu, %5.85'nin (n = 34) Akdeniz, %5.51'inin (n = 32) Karadeniz, %5.34'ünün (n = 31) Doğu Anadolu ve %2.06'sının (n = 12) Güneydoğu Anadolu Bölgelerinden olduğu gözlenmiştir. DFA örnekleminde de öğretmenlerin tamamı ortaokul düzeyindedir. Öğretmen uzmanlıkları ise AFA örneklemini ile bire bir örtüşecek şekilde projeye başvuran öğretmenler arasından seçkisiz olarak yapıldığından AFA ve DFA örneklemlerinin uzman dağılımları aynıdır.

## 2.3. İstatistiksel Analiz

İlk olarak betimsel analizler ve daha sonra madde analizleri SPSS 25.0 (IBM, 2017) paket programı ile yürütülmüştür. Açımlayıcı yöntem olarak varimax döndürme tekniği ile uygulanan temel bileşenler analizi; bileşen sayısı belirleme yöntemi olarak kullanılan paralel analiz ve MAP testi ise FACTOR 10.10 (Ferrando & Lorenzo-Seva, 2017) programı ile gerçekleştirilmiştir. Son olarak doğrulayıcı faktör analizi ise LISREL 8.80 (Jöreskog & Sörbom, 2006) programı kullanılarak yapılmıştır.

## 3. Bulgular

### **Bağlam Değerlendirme Ölçeği için Açımlayıcı Faktör Analizi ve Güvenirlik Çalışması**

İlk aşamada açımlayıcı faktör analizi için 581 katılımcıdan oluşan uzmanlar ile çalışılmış, temel bileşenler analizi yapabilmek için gerekli varsayımlar incelenmiş ve uygun sonuçlara ulaşılmıştır;  $KMO = .77$ , Bartlett Yakınsak  $\chi^2 = 1237.36$ ,  $p < .001$ ,  $Determinant = .12$ . Özdeğerlere bakıldığında ilki 3.12; ikincisi 1.00 olarak hesaplanan iki boyut tespit edilmiş; rotasyon öncesi matriste yalnızca madde 3'ün her iki bileşene de eşit faktör yüküyle (.66) yerleştiği tespit edilmiş, rotasyon işleminden sonra ise madde 3 de .87 faktör yüküyle birinci bileşene yerleşmiştir. Paralel analiz ve MAP testleri sonucunda önerilen boyut testleri de uyumlu olarak tek bileşenli bir yapıyı tespit edilmiştir. Tek faktörlü olduğu belirlenen bağlam değerlendirme ölçeği toplam varyansın %51.94'ünü açıklamaktadır.

Açımlayıcı faktör analizi örneklemini ile yapılan güvenirlik analizinde Cronbach alfa değeri .81 olarak hesaplanmış ve en düşük madde toplam puan korelasyonu .49 olarak belirlenmiştir. Temel bileşenler analizine ait faktör yükleri ve madde toplam puan korelasyonları Tablo 1'de gösterilmiştir.

**Tablo 1**

*Bağlam değerlendirme ölçeği açımlayıcı faktör analizi ve madde analizi bulguları*

| Madde                                                              | $\lambda$ | $r^*$ |
|--------------------------------------------------------------------|-----------|-------|
| 1. Bağlam sınıf düzeyine uygundur.                                 | .65       | .49   |
| 2. Soru bağlam gerçek bir problem durumunu yansıtmaktadır.         | .76       | .61   |
| 3. Bağlamda kullanılan görsel materyaller sorular için gereklidir. | .66       | .52   |
| 4. Bağlamda kullanılan senaryo/tablo/resim/şekil vb. uygundur.     | .74       | .63   |
| 5. Bağlam ilgi çekicidir.                                          | .77       | .62   |
| 6. Bağlam metni açık ve anlaşılırdır.                              | .73       | .57   |

$\lambda$  : Faktör yükü,  $r$  : Madde toplam puan korelasyonu

**Bağlam Değerlendirme Ölçeği için Doğrulatoryı Faktör Analizi**

Bağlam değerlendirme ölçeği için doğrulatoryı faktör analizi yine 581 katılımcıdan oluşan ikinci örneklem ile yürütülmüştür. Doğrulatoryı faktör analizi sonucunda 6 maddelik ölçeğin kabul edilebilir düzeylerde uyum gösterdiği saptanmıştır; *Satorra Bentler*  $\chi^2 = 23.27$ ,  $sd = 9$ ,  $p < .001$ ,  $\chi^2/sd = 2.58$ ,  $RMSEA = .09$  (%90 Güven Aralıkları: .05-.13),  $SRMR = .05$ ,  $NFI = .99$ ,  $CFI = .99$ ,  $GFI = .97$ . Standart faktör yükleri incelendiğinde en düşük faktör yükünün .54 olduğu gözlenmiş, doğrulatoryı faktör analizine ait değerler Tablo 2’de ve yol diyagramı Şekil 1’de gösterilmiştir.

**Tablo 2**

*Bağlam değerlendirme ölçeği doğrulatoryı faktör analizi bulguları*

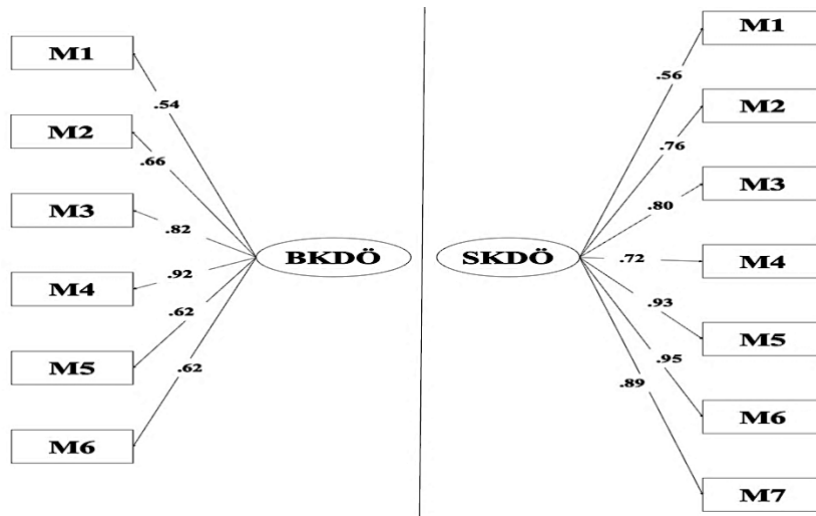
|         | $L$  | $\lambda$ | $S. Hata$ | $R^2$ | $t\text{-Değeri}$ |
|---------|------|-----------|-----------|-------|-------------------|
| Madde 1 | 1.00 | .54       | .70       | .30   |                   |
| Madde 2 | 1.10 | .66       | .56       | .44   | 11.66*            |
| Madde 3 | 2.24 | .82       | .32       | .68   | 8.20*             |
| Madde 4 | 1.95 | .92       | .13       | .87   | 8.41*             |
| Madde 5 | 0.60 | .62       | .60       | .40   | 8.26*             |
| Madde 6 | 1.06 | .62       | .62       | .38   | 14.00*            |

$L$ : Madde toplam puan korelasyonu,  $\lambda$  : Faktör yükü

\* $p < .05$

**Şekil 1**

*Bağlam Kalitesi (BDKÖ) ve Soru Kalitesi Değerlendirme Ölçeklerine (SDKÖ) Ait Doğrulatoryı Faktör Analizi Yol Diyagramı*



### **Soru Kalitesi Değerlendirme Ölçeği için Açımlayıcı Faktör Analizi ve Güvenirlik**

Soru kalitesi değerlendirme ölçeğinin 7 maddesi için de açımlayıcı faktör analizi bağlam kalitesi değerlendirme ölçeğinin açımlayıcı faktör analizi örnekleme ile yürütülmüş ve veri özellikleri analize uygun bulunmuştur;  $KMO = .84$ , Bartlett Yakınsak  $\chi^2 = 1603.88$ ,  $p < .001$ ,  $Determinant = .06$ . Özdeğerler hesaplamalarına göre özdeğeri 3.63 olan tek bir bileşen saptanmış; paralel analiz ve MAP testi de bir bileşen önermiştir. Tek faktörlü bir yapı gösterdiği belirlenen soru kalitesi değerlendirme ölçeğinin toplam varyansın %52'sini açıkladığı saptanmıştır. Soru kalitesi değerlendirme ölçeğinin en düşük faktör yükünün .65, en yüksek faktör yükü değerinin de .77 olduğu tespit edilmiştir (Tablo 3).

Aynı örnekleme yürütülen güvenirlik analizlerinde Cronbach alfa değeri .84 olarak belirlenirken; en düşük madde toplam puan korelasyonunun ise .43 olduğu görülmüştür. Açımlayıcı faktör analizi ve madde analizi bulguları Tablo 3'te gösterilmiştir.

**Tablo 3**

*Soru değerlendirme ölçeği açımlayıcı faktör analizi ve madde analizi bulguları*

|                                                                                  | $\lambda$ | $r^*$ |
|----------------------------------------------------------------------------------|-----------|-------|
| 1. Soru, üst düzey düşünme becerilerini ölçmektedir.                             | .65       | .43   |
| 2. Tercih edilen soru formatı uygundur.                                          | .76       | .67   |
| 3. Sorunun tek bir doğru cevabı (veya en doğru cevap) vardır.                    | .66       | .50   |
| 4. Sorunun çözümünde beklenen becerilerin dışında bir beceri kullanılmamaktadır. | .74       | .49   |
| 5. Verilen bilgiler soru çözümünün yapılabilmesi için yeterlidir.                | .77       | .69   |
| 6. Soru açık ve anlaşılırdır.                                                    | .73       | .75   |
| 7. Soru dil ve anlatım açısından kurallara ve sınıf düzeyine uygundur.           | .70       | .64   |

$\lambda$  : Faktör yükü,  $r$  : Madde toplam puan korelasyonu

### **Soru Değerlendirme Ölçeği için Doğrulayıcı Faktör Analizi**

Bağlam değerlendirme ölçeğinin doğrulayıcı faktör analizlerindeki örneklem ile soru değerlendirme ölçeği için de doğrulayıcı faktör analizleri yürütülmüştür. Elde edilen uyum indekslerine göre modelin kabul edilebileceği görülmüştür;  $Satorra Bentler \chi^2 = 62.92$ ,  $sd = 14$ ,  $p < .001$ ,  $\chi^2/sd = 4.49$ ,  $RMSEA = .08$  (%90 Güven Aralıkları: .05-.14),  $SRMR = .03$ ,  $NFI = .99$ ,  $CFI = .99$ ,  $GFI = .96$ . DFA modelinin bulguları Tablo 4'te, modele ait yol diyagramı ise Şekil 1'de gösterilmiştir.

**Tablo 4**

*Soru Değerlendirme Ölçeği Doğrulayıcı Faktör Analizi Bulguları*

|         | $L$  | $\lambda$ | $S. Hata$ | $R^2$ | $t$ -Değeri |
|---------|------|-----------|-----------|-------|-------------|
| Madde 1 | 1.00 | .56       | .69       |       |             |
| Madde 2 | 1.98 | .76       | .43       | .31   | 12.15*      |
| Madde 3 | 3.22 | .80       | .36       | .57   | 9.59*       |
| Madde 4 | 1.50 | .72       | .48       | .64   | 9.27*       |
| Madde 5 | 3.17 | .93       | .13       | .52   | 10.52*      |
| Madde 6 | 2.93 | .95       | .10       | .87   | 10.09*      |
| Madde 7 | 2.59 | .89       | .31       | .90   | 10.25*      |

$L$ : Madde toplam puan korelasyonu,  $\lambda$  : Faktör yükü

\* $p < .05$

### **İki Ölçek Arasındaki İlişkinin Belirlenmesi**

Ölçeklerin yapı geçerliği ve güvenirlik çalışmalarından sonra açımlayıcı ve doğrulayıcı faktör analizi örneklemeleri birbirine eklenerek 1162 katılımcıdan oluşan örnekleme ölçeklere ait toplam puanlar hesaplanmıştır. Bağlam kalitesi

değerlendirme ölçeği için çarpıklık değeri -.64 ve basıklık değeri -.47 olarak belirlenirken soru kalitesi değerlendirme ölçeği için çarpıklık değerinin -.97 ve basıklık değerinin .24 olduğu görülmüştür. Ayrıca bağlam değerlendirme ölçeğine ait z değerleri -2.78 ile 1.88 arasında dağılırken soru değerlendirme ölçeğine ait z puanı aralığı -3.01 ile 2.58 olarak belirlenmiştir. Bu analizlerden hareketle normal dağılım varsayımı kabul edilmiştir. İki ölçeğin toplam puanları arasında yapılan Pearson korelasyonu sonucunda iki ölçek arasında istatistiksel olarak anlamlı orta kuvvette pozitif bir ilişkinin olduğu saptanmıştır;  $r = .57, p < .001$ .

#### 4. Sonuç ve Tartışma

Bu çalışma temelde uzmanların bağlam ve soruları değerlendirdiği durumlarda kullanılabilecek niceliksel bir ölçme aracının gerekliliği sebebiyle yürütülmüş ve bu bağlamda uygun ölçekler geliştirilmiştir. Bağlamlar ve bunlara bağlı soruların sınıf, konu ve öğrenci düzeyine uygunluklarının hakemler tarafından değerlendirilebilmesi amacıyla tek faktörlü iki ayrı ölçek geliştirilmiştir. Bağlam değerlendirme ölçeği, üst düzey düşünmeyi ölçen sorulara ya da diğer bir ifade ile yeni nesil sorulara ait ana gövde metnini oluşturan bağlamların uygunluğunun değerlendirilmesi için geliştirilmişken; soru değerlendirme ölçeği de bu bağlama bağlı soruların uygunluğunun değerlendirilmesi amacıyla geliştirilmiştir. Bağlam sayısı her zaman bir olmasına rağmen bağlama bağlı soruların sayısı değişkenlik göstermektedir. Bu nedenle örneğin 2 soruya sahip bir bağlam değerlendirilirken bağlam değerlendirme ölçeğinin yalnız bir kere; soru değerlendirme ölçeğinin ise iki kere kullanılması gerekmektedir. Bağlamların bir bütün halinde kullanılabilmesi için kendi kalitesi kadar, ilişkili soruların kalitesinin de yüksek olması gerektiğinden böyle bir durumda üç değerlendirmeden gelen ölçek puanlarının ortalamasının alınması önerilmektedir.

Soru kalitelerinin ve bağlam kalitelerinin ayrı ayrı ve bir arada ele alınmasının önemi literatürde de görülmektedir (Halpern, 2013; Paul & Elder, 2008). Hem bağlamların hem soruların; transfer, analitik çözümleme ve eleştirel düşünme gibi üst düzey düşünme becerilerine ait özellikleri ölçmesinin yanı sıra sınıfa uygun olması, anlaşılabilir olması gibi göstergeler de bağlamların ve soruların kalitesinin parçalarıdır. Eğitimde üst düzey düşünme becerilerin etkili bir şekilde ölçülmesinin önemi vurgulanmaktadır (Biggs & Tang, 2011; Brown, Collins & Duguid, 1989; Zimmerman, 2002).

En temel üst düzey düşünme becerisi olan transfer becerilerinin öğrencilerin bilgiyi farklı bağlamlarda uygulamasındaki rolü eğitimde önemli bir yer tutmaktadır. Karmaşık soruları veya sorunları parçalarına ayırmayı ve kanıtları sistematik olarak düzenleyip değerlendirmeyi içeren analitik düşünme becerisi de bağlamlar içerisinde temsil edilmesi gereken bir beceridir. Jonassen (2000) tarafından yapılan çalışmada becerilerin bir arada sorularda ele alınmasının gerekliliği gösterilmiştir. Problem temelli öğrenme süreçlerinin analitik çözümleme becerisini artırdığı ve bunun ancak bağlama dayalı sorularla mümkün olduğu bu çalışmada ortaya konmuştur. Son olarak en üst düzey beceri olan eleştirel düşünme becerisi, argümanları sağlam bir şekilde değerlendirip sentezleyebilmeyi ifade etmektedir (Facione, 1990). Elder ve Paul (2008) de bağlamlarda tüm bu üç becerinin ele alınmasının önemini vurgulamıştır.

Öte yandan bağlamların ve soruların kalitesi de üst düzey düşünme becerilerinin ölçümünde önemli bir rol oynamaktadır. Herrington ve Oliber (2000); Brown ve diğ. (1987) gibi birçok araştırmacı eğitimde beceri kazanımlarını değerlendirmek için kullanılan bağlam ve soruların sınıf ve beceri düzeyine ilgisinin ve özgünlüğünün önemini vurgulamıştır. İyi hazırlanmış bağlamların beceriyi ölçerken aynı zamanda üst düzey düşünme beceri seviyesini de artırdığı gözlenmiştir. Sınavlarda, testlerde ya da ölçme araçlarında kullanılacak bağlamların konularla alakalı, açık, anlaşılabilir olması değerlendirmenin geçerlik ve güvenirliği için oldukça önemlidir. Bağlama bağlı soruların tasarımı da oldukça önemlidir. Öğrenme hedeflerine uygun olarak belirlenmiş ve üst düzey düşünme becerilerini destekleyen

sorular; öğrenci becerilerinin etkili bir şekilde değerlendirilmesi bakımından önemlidir (Biggs & Tang, 2011; Bloom, 1956). Beceri kazanımlarının ölçüldüğü araçlarda öğrenciler bilgiyi analiz etmeye, sentezlemeye, yeniden üretmeye ve değerlendirmeye teşvik edilmelidir (Perkins & Salomon, 1988).

Hem bağlam hem de soru değerlendirme ölçeklerinin geçerli ve güvenilir sonuçlar verdiği gözlenmiştir. Açıklayıcı faktör analizi ile tek boyutlu yapıda olduğu tespit edilen bağlam ve soru kalitesi değerlendirme ölçeklerinin doğrulayıcı faktör analizi bulguları da ölçeklerin yapısal bakımdan geçerli olduğunu ortaya koymuştur. Bağlam kalitesi değerlendirme ölçeğinin Cronbach alfa iç tutarlık güvenirlik katsayısı ise .81; soru kalitesi değerlendirme ölçeğinin alfa değeri .84 olarak hesaplanmış ve güvenilir oldukları ortaya konmuştur. Her iki ölçeğin de uzmanlar tarafından bağlam değerlendirilirken kullanılabileceği ortaya konmuştur. Likert formatında 5 dereceli olarak değerlendirilen ölçeklerden genel toplam puan elde edilmemektedir. Ölçeklerin genel toplam puan elde edilememesinin sebebi; aynı bağlama bağlı olarak sorulan soruların standartının olmamasıdır. Bazı bağamlar için yalnızca bir soru kullanılırken bazı bağamlar için birden çok soru kullanılmaktadır. Bu nedenle araştırmacılara ölçeklerden elde edilen puanları değerlendirirken elde edilen ortalama puanlar üzerinden örtük sınıf analizi kullanarak tepki olasılığı yüksek bağlam ve soruları seçmeleri önerilmektedir. Diğer bir alternatif ise bağlam ve soru kalitesi değerlendirme ölçeklerinden gelen toplam puanların ortalaması elde edilerek 3 (Likert tipi ölçeklerde elde edilebilecek ortalama puan) puanı aşan bağlam ve soruların kaliteli sorular olarak değerlendirilip ilgili testlere dahil edilmesidir. Bir bütün olarak bağlam ve soruların belirlenmesinde ise bağlam kalitesi değerlendirme ölçeğinden gelen ortalama puan ile sorulara ait kalite ölçümlerinden gelen ortalamaların ortalaması alınarak genelde yine 3 puanı aşan bağamların kullanılmasıdır.

Üst düzey düşünme becerilerinin ölçülmek istendiği ölçümlerde uzmanların bağlam seçerken bu ölçekleri kullanabileceği belirlenmiştir. Bulgular, aracın test, sınav vb. ölçümler oluşturmalarında yardımcı olacaktır. Ancak araştırma kapsamında yalnızca iç-tutarlık güvenirlik analizleri ve yapı geçerliği analizleri kullanılmıştır. Benzer bir yapıyı ölçen kriter bir ölçek bulunmadığında kriter ilişkili geçerlik çalışması yapılamamıştır. Öte yandan test-tekrar test güvenirliği de bağlam ve soru kalitesinin doğasından kaynaklı olarak uygun bulunmamıştır. Araştırmada sadece ilgili TÜBİTAK projesine katılan ve bağlam değerlendirmeye gönüllü olan öğretmenlerle çalışılmıştır.

Sonuç olarak, bu araştırmada üst düzey düşünme becerilerini değerlendirme ve geliştirmede yüksek ve düşük kaliteli bağlam ve soruların ayırt edilmesine katkı sağlayacak bir ölçüm aracı geliştirilmiştir. Ölçüm aracının güvenirlik ve geçerliği gösterilmiştir. Kuramlarla uyumlu bir ölçek ortaya konmuştur.

## **Çıkar Çatışması Bildirimi**

Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

## **Etik Beyan**

Bu araştırma, insan katılımcılardan anket yoluyla veri toplamayı içermektedir. Çalışma, Ege Üniversitesi Sosyal ve Beşerî Bilimler Etik Kurulu tarafından onaylanmıştır (Etik Kurul Onay Numarası: 721 / Tarih: 02.12.2020). Tüm katılımcılardan çevrim içi onam alınmıştır. Araştırma süreci, 1964 Helsinki Bildirgesi ilkelerine uygun şekilde yürütülmüştür.

## **Üretken Yapay Zekâ Kullanımına Dair Beyan**

Bu çalışmanın hazırlanma sürecinde üretken yapay zekâ (ChatGPT, Bard, Claude, Copilot vb.) araçları kullanılmamıştır. Tüm metin, analiz ve içerik, yazarlar tarafından insan katkısıyla üretilmiştir.

## Yazar Katkıları

Tüm yazarlar, çalışmanın kuramsal çerçevesinin oluşturulması ve araştırmanın tasarlanmasına katkı sağlamıştır. Veri toplama sürecinin planlanması, veri toplama ve analiz süreçleri (Yılmaz Orhun Gürlük) tarafından yürütülmüştür. Tartışmaya (Yılmaz Orhun Gürlük ve Mediha Korkmaz ve Tahsin Oğuz Başokçu) katkı sağlamıştır. Makalenin ilk taslağı (Yılmaz Orhun Gürlük) tarafından yazılmış, tüm versiyonlar tüm yazarlar tarafından gözden geçirilmiştir. Tüm yazarlar, makalenin son hâlini okuyup onaylamıştır.

## Teşekkür

Bu yayın için veri toplanması sürecindeki katkılarından dolayı TÜBİTAK ve İzmir İl Milli Eğitim Müdürlüğüne teşekkürlerimizi sunarız.

## Kaynakça

- Alaimi, M., Law, E., Pantasdo, K. D., Oudeyer, P.-Y., & Sauzeon, H. (2020). Pedagogical agents for fostering question-asking skills in children. *arXiv preprint arXiv*. <https://doi.org/10.48550/arXiv.2004.03472>
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612-637. <http://doi.org/10.1037/0033-2909.128.4.612>
- Baysal, E. A., & Ocak, G. (2022). University Students' Cognitive Bias in the Context of Their Analytical Thinking Skills: A Reliability and Validity Study. *International Journal of Progressive Education*, 18(3), 205–225. <https://doi.org/10.29329/ijpe.2022.439.14>
- Beno, P., & Kovacova, B. (2020). Structured, Analytical and Critical Thinking in the Educational Process of Future Teachers. *Acta Educationis Generalis*, 10(3), 111–118. <https://doi.org/10.2478/atd-2020-0024>
- Biggs, J., & Tang, C. (2011). *Teaching for Quality Learning at University* (Eds. 4). Open University Press.
- Bloom, B. S. (1956). Taxonomy of Educational Objectives, *Handbook I: The Cognitive Domain*. David McKay Co Inc.
- Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, 24(1), 61-100.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. National Academy Press.
- Brown, A. L. (1987). Metacognition, executive control, self-regulation , and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.1), *Metacognition, Motivation, and Understanding*, 65-116.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32-42. <https://doi.org/10.3102/0013189X018001032>
- Butterfield, E. C., & Nelson, G. D. (1989). Theory and practice of teaching for transfer. *Educational Technology Research and Development*, 37(3), 5-38. <http://doi.org/10.1007/BF02298274>
- Butterfield, E. C., Nelson, T. O., & Peck, V. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology*, 24(5), 654–663. <http://doi.org/10.1037/0012-1649.24.5.654>

- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Educational Technology Research and Development*, 37, 5-38. [http://doi.org/10.1207/s15516709cog1302\\_1](http://doi.org/10.1207/s15516709cog1302_1)
- Elballah, K., Alkhalifah, N., Alomari, A. & Alghamdi, A. (2024). Enhancing cognitive dimensions in gifted students through future problem-solving enrichment programs. *Discover Sustainability*, 5, 248. <https://doi.org/10.1007/s43621-024-00470-5>
- Elder, L. & Paul, R. (2008). Critical thinking: the nuts and bolts of education. *Optometric Education*, 33(3), 88-91.
- Ennis, R. H. (1996). *Critical Thinking*. Prentice Hall.
- Facione, P. A. (1990). Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. *The Delphi Report*. <http://doi.org/10.1037/0022-0663.91.2.260>
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10), 906-911. <https://doi.org/10.1037/0003-066X.34.10.906>
- Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52, 67-86. <http://doi.org/10.1007/BF02504676>
- Halpern, D. F. (2013). *Thought and Knowledge: An Introduction to Critical Thinking* (5th ed.). Psychology Press.
- Hasanah, S., & Retnawati, H. (2022). Assessment of contextual learning in mathematics education. *AIP Conference Proceedings*, 2575(1), 040018. <https://doi.org/10.1063/5.0111142>
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81-112. <http://doi.org/10.3102/003465430298487>
- Herrington, J., & Oliver, R. (2000). An instructional design framework for authentic learning environments. *Educational Technology Research and Development*, 4, 23-48. <http://doi.org/10.1007/BF02319856>
- Hidayat, R., Nugroho, I., Zainuddin, Z., & Ingai, T. A. (2024). A systematic review of analytical thinking skills in STEM education settings. *Information and Learning Sciences*, 125(7/8), 565–586. <https://doi.org/10.1108/ILS-06-2023-0070>
- IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48, 63-85. <http://doi.org/10.1007/BF02300500>
- Jöreskog, K.G. & Sörbom, D. (2006). LISREL 8.80 for Windows [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.
- Krell, M., Kind, V., & Krüger, D. (2024). Validation of the Scientific Reasoning Competencies Instrument: Relationships with Epistemological Beliefs and Analytical Thinking. *International Journal of Science and Mathematics Education*. <https://doi.org/10.1007/s10763-024-10482-2>
- Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, 28(2), 16-25. <http://doi.org/10.3102/0013189X028002016>

- Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.
- Ferrando & Lorenzo-Seva (2017). Program FACTOR at 10: Origins, development and future directions. *Psicothema*, 29(2), 236-240. <http://doi.org/10.7334/psicothema2016.304>.
- Mayer, R. E. (2002). Rote versus meaningful learning. *Theory into Practice*, 41(4), 226-232. [http://doi.org/10.1207/s15430421tip4104\\_4](http://doi.org/10.1207/s15430421tip4104_4)
- Paris, S. G., & Winograd, P. (1990). How metacognition can promote academic learning and instruction. In B. F. Jones & L. Idol (Eds.1), *Dimensions of Thinking and Cognitive Instruction*, 15-51.
- Paul, R., & Elder, L. (2008). *The Miniature Guide to Critical Thinking: Concepts and Tools*. Foundation for Critical Thinking Press.
- Perkins, D. N., & Salomon, G. (1988). Teaching for Transfer. *Educational Leadership*, 46(1), 22-32.
- Rutumalessy, M., Sudyana, I. N., Azis, A. A., Suharyatun, S., & Suroso, A. (2023). The Implementation of Contextual Learning and Teaching Method in Improving Students Learning Achievement. *Journal on Education*, 5(4), 16262–16267. <https://jonedu.org/index.php/joe/article/view/2771>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144. <http://doi.org/10.1007/BF00117714>
- Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice*, 41(4), 219-225. [http://doi.org/10.1207/s15430421tip4104\\_3](http://doi.org/10.1207/s15430421tip4104_3)
- Schraw, G., & Moshman, D. (1995). Metacognitive Theories. *Educational Psychology Review*, 7(4), 351-371. <http://doi.org/10.1007/BF02212307>
- Spiro, R. J., Coulson, R. L., Feltovich, P. J., & Anderson, D. K. (1988). Cognitive Flexibility Theory: Advanced Knowledge Acquisition in Ill-Structured Domains. In V. Patel (Ed.1), *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*, 375-383.
- Sternberg, R. J. (1986). *Critical Thinking: Its Nature, Measurement, and Improvement*. National Institute of Education.
- Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of Educational Psychology*, 82(2), 306-314. <http://doi.org/10.1037/0022-0663.82.2.306>
- Tobias, S., & Everson, H. T. (2000). Assessing metacognitive knowledge monitoring. In G. Schraw & J. Impara (Eds.2), *Issues in the Measurement of Metacognition*, 147-222.
- Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, 1, 3-14. <http://doi.org/10.1007/s11409-006-6893-0>
- Wang, Y., & Chen, L. (2023). Enhancing student's computational thinking skills with student-generated questions strategy in a game-based learning platform. *Computers & Education*, 200. <https://doi.org/10.1016/j.compedu.2023.104794>

- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds. 1), *Metacognition in Educational Theory and Practice*, 277-304.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100. <http://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Yu, X., Wang, T.-Y., & Yuizono, T. (2023). Creativity development through questioning activity in second language education. *Frontiers in Education*, 8, 1178655. <https://doi.org/10.3389/feduc.2023.1178655>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64-70. [http://doi.org/10.1207/s15430421tip4102\\_2](http://doi.org/10.1207/s15430421tip4102_2)

## Extended Abstract

### Introduction

In modern education, measurement methods determined based on contexts are generally based on higher-order thinking skills. High-order thinking skills are generally grouped in three classes as transfer skill, which can be defined as the ability to express acquired or existing information in different formats; analytical solution skill, which can be defined as the ability to synthesize methods by finding new solutions; and finally, critical thinking skill, which can be defined as the ability to question and change information or methods. While the contexts in which these skills are measured constitute the main body of the question or questions; context-based questions are the parts in which the actual knowledge and skills are measured. Both the context and the questions should focus on the relevant skill, be clear and understandable, and be appropriate for the grade level. Such features determine the quality of both the contexts and the questions. Context and questions are interdependent with each other. The fact that the context is good and the questions are bad or the context is bad and the questions are good causes the measurement as a whole to be inappropriate. This quality assessment is usually carried out qualitatively by experts. This study aims to develop quantitative measurement tools that experts can use when evaluating the context and questions. Within the scope of the study, both context quality and question quality evaluation scales were developed.

### Methodology

The research was conducted within the scope of the doctoral thesis of the lead author of this article and within the scope of the TÜBİTAK 1001 scientific research project numbered 120K850. Teachers from different regions of Turkey were reached in the research. In order to prevent problems that may arise due to sample size differences, both exploratory and confirmatory factor analyses were conducted with samples consisting of 581 participants. First, exploratory factor analysis, item analysis and confirmatory factor analysis were conducted for the context quality assessment scale and then the same analyses were repeated for the question quality assessment scale. Descriptive statistics, item analysis and exploratory factor analysis findings were conducted with the SPSS 25.0 package program. Confirmatory factor analyses were conducted with LISREL 8.80. In addition, the FACTOR program was used to provide exploratory factor analysis, to see the parallel test, and to see the MAP test.

### Results

Experts consisting of 581 participants were studied for EFA for the context quality assessment scale, the necessary assumptions were examined to be able to perform principal component analysis and appropriate results were reached;  $KMO = .77$ , *Bartlett Convergent*  $\chi^2 = 1237.36$ ,  $p < .001$ , *Determinant* = .12. When the eigenvalues were examined, two dimensions were determined, the first of which was calculated as 3.12 and the second as 1.00; it was determined that item 3 was placed in both components with equal factor loading (.66) in the matrix before rotation, and after the rotation, item 3 was placed in the first component with a factor loading of .87. As a result of the parallel analysis and MAP tests, a single-component structure was determined in accordance with the suggested dimension tests. The context assessment scale, which was determined to be uni-dimensional, explains 51.94% of total variance. In the reliability analysis conducted with the exploratory factor analysis sample, alpha value was calculated as .81 and the lowest item-total score correlation was determined as .49. As a result of CFA, it was determined that the 6-item scale showed acceptable levels of fit; *Satorra Bentler*  $\chi^2 = 23.27$ ,  $df = 9$ ,  $p < .001$ ,  $\chi^2/df = 2.58$ ,  $RMSEA = .09$  (90% CI: .05-.13),

$SRMR = .05$ ,  $NFI = .99$ ,  $CFI = .99$ ,  $GFI = .97$ . When the standard factor loadings were examined, it was observed that the lowest factor loading was .54.

The same analyses were then repeated for the question quality assessment scale. EFA analysis for 7 items of the question quality assessment scale was conducted with the exploratory factor analysis sample of the context quality assessment scale and the data characteristics were found suitable for the analysis;  $KMO = .84$ , *Bartlett Convergent*  $\chi^2 = 1603.88$ ,  $p < .001$ , *Determinant* = .06. According to the eigenvalue calculations, a single component with an eigenvalue of 3.63 was detected; parallel analysis and MAP test also suggested one component. It was found that the question quality assessment scale explained 52% of total variance. It was found that the lowest factor loading of the question quality assessment scale was .65 and the highest factor loading was .77. In the reliability analyses conducted on the same sample, alpha value was determined as .84; the lowest item-total score correlation was seen as .43. According to the fit indices obtained in CFA for the question assessment scale, it was seen that the model could be accepted; *Satorra Bentler*  $\chi^2 = 62.92$ ,  $df = 14$ ,  $p < .001$ ,  $\chi^2/df = 4.49$ ,  $RMSEA = .08$  (90% CI: .05-.14),  $SRMR = .03$ ,  $NFI = .99$ ,  $CFI = .99$ ,  $GFI = .96$ .

## Conclusion and Discussion

This study focuses on developing a quantitative measurement tool that can be used in situations where experts evaluate contexts and questions. Effectively measuring higher-order thinking skills (transfer, analytical solution and critical thinking) is of great importance in education. The quality of context and questions plays an important role in measuring these skills. Well-prepared contexts increase higher-order thinking skills and for evaluation validity, contexts must be relevant to the topics, clear and understandable. It was observed that the scales developed in this study were valid and reliable, and the construct validity and internal consistency reliability of the context and question quality evaluation scales were high. It was suggested that the scales be evaluated based on the average score instead of the overall total score.

Although the number of contexts is always one, the number of questions related to the context varies. For example, when evaluating a context with 2 questions, the context assessment scale should be used only once, and the question assessment scale should be used twice. Since the quality of the related questions as well as the quality of the contexts should be high in order to be used as a whole, it is recommended to average the scale scores from the three evaluations in such a case.

As a result, this study presented measurement tools that distinguishes high- and low-quality contexts and questions to be used in evaluating and developing higher-order thinking skills. The reliability and validity of the measurement tools have been proven and it has been determined that experts can use the scales in education and other measurement tools.