



PISA 2015 Okuma Becerisi Maddelerinin Güçlük İndeksini Etkileyen Madde Özelliklerinin İncelenmesi

*Sinem DEMİRKOL**
*Merve AYVALLI KARAGÖZ***

Öz

Bu çalışmanın amacı okuma becerisi maddelerinin güçlük indeksini etkileyen madde özelliklerini belirlemektir. Bu amaç doğrultusunda madde formatı, madde bilişsel alan düzeyi ve bu iki değişkene ait etkileşimin madde güçlüğü üzerindeki etkileri incelenmiştir. Araştırmanın çalışma grubunu PISA 2015 Türkiye uygulamasında okuma becerisi alt testine yanıt veren 2418 öğrenci oluşturmaktadır. Çalışmanın analizleri çok seviyeli bir yöntem olan Açıklayıcı MTK modelleri ile yürütülmüştür. Elde edilen sonuçlar açık uçlu maddelerin çoktan seçmeli maddelere göre, anlama ve yorumlama bilişsel alanında yer alan maddelerin ise bilgi ve değerlendirme basamağında yer alan maddelere göre anlamlı derecede daha zor olduğunu göstermektedir. Madde formatı ve madde bilişsel alan kesişimi incelendiğinde ise, bilişsel alanı anlama ve yorumlama olan maddelerinin açık uçlu sorulmasının maddeleri kolaylaştıracağı, bilgi basamağında yer alan maddelerin ise açık uçlu sorulmasının maddeleri zorlaştıracağı saptanmıştır.

Anahtar Kelimeler: Madde formatı, bilişsel alan, açıklayıcı madde tepki kuramı, PISA 2015

Investigation of Item Properties Affecting the Difficulty Index of PISA 2015 Reading Literacy Items

Abstract

Makaleni The aim of this research is to investigate the item properties that affect the parameters of the reading literacy items. For this purpose, item format, cognitive domain and the effects of the interaction of these two variables on item difficulty are examined. The study group of the research consists of 2418 students who responded to the reading subtest in the PISA 2015 Turkey sample. The analyzes of the study are carried out with Explanatory IRT models, which is a multi-level method. According to the results, open-response items are significantly more difficult than multiple-choice items, and the items in the integrate cognitive domain are significantly more difficult than the items in the access and evaluate domain. In addition, constructed-response items are more suitable for measuring meta-cognitive domains, while selected-items are better for measuring achievements in sub-cognitive domains.

Keywords: Item format, cognitive domain, explanatory IRT, PISA 2015

Giriş

Eğitim sisteminin vazgeçilmez bir ihtiyacı olan ölçme değerlendirme işlemlerinin verimli ve etkin bir şekilde yürütülebilmesi için en sık kullanılan ölçme araçlarından biri bireylerin bilişsel, duyuşsal ve devinişsel özelliklerini belirlemek için kullanılan testlerdir (Crocker ve Algina, 1986). Testin türü, kullanım amacı, ölçülmesi planlanan davranış özellikleri ve testte kullanılan madde tipi testin yapısını

* Dr., Ordu Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Ordu, sinemdemirkol@odu.edu.tr, ORCID: orcid.org/0000-0002-9526-6156

** Dr., Akdeniz Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Antalya, merveyayvalli@akdeniz.edu.tr, ORCID: orcid.org/0000-0002-7301-0096

oluşturan temel özelliklerdir (Demir, 2010). Ölçmenin yapılma amacı, kullanılacak testin ve testte yer alan maddelerin özelliğini de etkilemektedir. Crocker ve Algina'ya (1986) göre, neyin nasıl ölçüleceğine karar verirken uygun madde formatını seçilmesi ve bu formatın öğrenciler için uygun olup olmadığına karar verilmesi oldukça önemlidir. Madde formatının seçiminde öğrenme çıktıları, öğrencilerin hazırbulunuşluk düzeyleri ve yaşları, değerlendirmenin amacı, ölçülmesi hedeflenen içeriğin/yapının özelliği ve testi hazırlayanların deneyimi gibi farklı değişkenler göz önüne alınmalıdır (Phipps ve Brackbill, 2009).

Madde formatı genel olarak öğrencilerin cevap şekline göre seçmeli-cevaplı ve yapılandırılmış-cevaplı olarak sınıflandırılmaktadır (Osterlind, 2013). Öğrencinin verilen seçenekler arasından bir ya da birden fazla seçim yapmasına izin veren çoktan seçmeli maddeler, doğru/yanlış maddeleri, evet/hayır maddeleri ve uygun resmin/grafiğin seçilmesi gibi maddeler seçmeli-cevaplı (selecte-response) maddeler olarak adlandırılır (Jennings ve Bush, 2006). Kısa cevaplı maddeler, kompozisyonlar, cevabın yazılması istenen maddeler, tamamlama maddeleri gibi cevabın öğrenciler tarafından oluşturulması istenen maddeler ve güzel sanatlar alanlarında ortaya çıkan eserler de yapılandırılmış-cevaplı (constructed-response) madde formatı başlığında ele alınır (Pollack, Rock ve Jenkins, 1992).

Ülkemizde en sık kullanılan madde formatı seçmeli maddeler (çoktan seçmeli, doğru/yanlış, evet/hayır), seçmeli madde formatında da en sık kullanılan madde formatı ise çoktan seçmeli maddelerdir. Çoktan seçmeli maddeler, doğru cevabın testi hazırlayan tarafından verildiği ve öğrencilerden doğru cevabın çeldiriciler arasından seçilmesinin istendiği madde türüdür. Çoktan seçmeli maddelerin objektif bir şekilde puanlanabilmesi (Walstad, 1998), madde analizlerinin daha kolay yapılabilmesi (Bible, Simkin ve Kuechler, 2008), geniş ölçekli testlerde hızlı ve etkili bir şekilde veri toplanabilmesi (Walstad ve Becker, 1994), hızlı cevaplanabildikleri için daha fazla maddenin sorulabilmesi ve dolayısıyla hazırlanan testin kapsam geçerliğinin ve duyarlılığın yüksek olması (Bennett, Ward, Rock ve LaHart, 1990) gibi üstün yanları mevcuttur. Bu avantajlarının yanında öğrencilerin tahmin yoluyla veya çeldiricileri eleyerek doğru cevaba ulaşmaları (Bush, 2001), dolayısıyla testin güvenilirliğinin daha düşük olması (Thawabieh, 2016), bilişsel düzeyin üst basamaklarında yer alan davranışları ölçmeye elverişli olmamaları (Bennett vd., 1990), hazırlanmasının uzmanlık ve uzun zaman gerektirmesi (Brown, Bull ve Pendlebury, 2013), öğrencilerin kısmi bilgilerinin göz ardı edilmesi (Ben-Simon, Budescu ve Nevo, 1997) gibi dezavantajları da bulunmaktadır.

Sık kullanılan bir başka madde formatı da yapılandırılmış-cevaplı maddelerdir. Bu madde türünde doğru cevap öğrenci tarafından organize edilerek sunulur (Geer, 1988). Yapılandırılmış cevaplı maddeler, öğrencilerin uygulama, değerlendirme ve yaratıcı düşünme gibi üst bilişsel becerilerini ölçmede daha etkilidir (Bacon, 2003; Badger ve Thomas, 1992; Hancock, 1994). Aynı zamanda bu madde formatı ile öğrenme çıktılarının yanında, öğrenme süreci de etkili bir şekilde ölçülebilir (Lee, Liu ve Linn, 2011). Maddelerin hazırlanması kolay ve hızlıdır (Badger ve Thomas, 1992). Çoktan seçmeli maddelerin dezavantajlarından biri olan tahmin ve çeldiricileri eleyerek doğru cevabı bulma olasılığı bu madde türünde oldukça düşüktür (Ruch ve Staddord, 1925). Kısmi puanlamaya uygun olduğundan tam olmayan öğrenmeler belirlenebilir. Bu gibi avantajlarının yanında uygulamasının ve puanlamasının zaman alması (Wainer ve Thissen, 1993), fazla maddenin sorulamamasından dolayı testin kapsam geçerliğini sağlamanın güç olması (Ventouras, Triantis, Tsiakas ve Stergiopoulos, 2010), derecelendirme anahtarı kullanılsa bile objektif puanlamanın zor olması (Powell ve Gillespie, 1990), yazma becerisi, yazı güzelliği, noktalama işaretlerinin doğru kullanımı gibi ölçmenin hedefi olmayan davranışların puanlamaya karışma ihtimali (Zeidner, 1987) gibi dezavantajları da bulunmaktadır.

Madde formatının yanında bir başka önemli madde özelliği de maddenin bilişsel alan düzeyidir. Öğrenme farklı zihinsel süreçlerin birleşiminden oluşmaktadır. Dolayısıyla hazırlanan testin, farklı bilişsel alanlarda maddeler içermesi, öğrencilerin hangi zihinsel süreçleri gerçekleştirebildiğini ortaya koyması açısından oldukça önemlidir. Ayrıca öğrencilerin karşılaştığı sorunların öğrenme ve zihinsel süreçleri geliştirdiği, sadece alt basamakta yer alan maddelerin (bilgi gibi) öğrencileri ezbere yönelttiği ve öğrencilerin eleştirel ve yaratıcı düşünme becerilerini geliştirmedeği belirtilmiştir (Gardner, Tremblay ve Masgoret, 1997). Bu yüzden iyi hazırlanmış bir test öğrencinin hedef davranışa ulaşma düzeyini ölçmenin yanında düşünme ve sorgulama becerilerini geliştirecek özelliklere de sahip olmalıdır (Hancock, 1994).

PISA bilişsel alan testlerinde yer alan maddeler farklı bilişsel düzeyleri ölçecek şekilde hazırlanmıştır (OECD, 2017b). PISA okuma alanına ait maddeler hazırlanırken “bilgiye ulaşma”, “anlama”, “yorumlama”, “metnin içeriğini yansıtma ve değerlendirme”, “metnin biçimini yansıtma ve değerlendirme” olmak üzere 5 bilişsel yönün ele alınması planlanmıştır (OECD, 2017b). Fakat test hazırlama sürecinde tüm bilişsel alanların eşit şekilde temsil edilmesi güç olduğundan beş boyut birleştirilerek, bilgi basamağı “bilgiye ulaşma ve hatırlama (access and retrieve)”, bilgiyi bir araya getirme ve yorumlama basamağı “anlama ve yorumlama (integrate and interpret)”, metnin içeriği ve biçimini yansıtma ve değerlendirme basamağı ise “yansıtma ve değerlendirme (reflect and evaluate)” başlığı altında birleştirilmiştir. Bilgiye ulaşma ve hatırlama basamağında öğrencilerden mevcut bilgilerini hatırlamaları, metnin içerisinde yer alan bilgileri seçmeleri veya sıralamaları istenmektedir. Bu düzeyde öğrencinin kendinden bir şey katmasına gerek yoktur. Anlama ve yorumlama basamağında öğrencilerden metnin içerisindeki ilişkilere odaklanmaları, karşılaştırma yapmaları, elde edilen bilgileri birleştirmeleri ve yorum yapmaları beklenmektedir. Yansıtma ve değerlendirme basamağında ise öğrencilerden bakış açılarını ifade etmeleri ve metnin özünü kavrayıp değerlendirmeleri istenmektedir (OECD, 2017b).

Literatür incelendiğinde madde formatının ve bilişsel alan düzeyinin madde parametrelerini, test özelliklerini ve öğrenci performanslarını etkilediğini ortaya koyan birçok çalışma mevcuttur. İlhan, Boztunç Öztürk ve Şahin (2020) TIMSS 2015 Türkiye verilerini kullanarak madde formatı ve madde bilişsel alan düzeyinin madde güçlüğü üzerindeki etkisini iki aşamalı bir yöntemle incelemişlerdir. Elde edilen sonuçlara göre açık uçlu maddelerin çoktan seçmelilere göre daha zor olduğu, maddelerin bilişsel alan düzeyi arttıkça da öğrenciler tarafından daha güç algılandığı görülmüştür. Griffo (2011) Ulusal Eğitim İlerleme Değerlendirmesi (NAEP) verilerini kullanarak madde formatının okuduğunu anlama üzerindeki etkisini incelemiştir. Elde edilen sonuçlara göre öğrencilerin en zor algılandığı madde formatının açık uçlu, en kolay anladığı madde formatının ise çoktan seçmeli olduğu görülmüştür. Hancock (1994) yürütmüş olduğu çalışmasında çoktan seçmeli ve yapılandırılmış cevap formatında hazırlanan maddelerin farklı bilişsel düzeyleri ne ölçüde ölçtüğünü incelemiş, alt bilişsel düzeyleri ölçmek için hazırlanan farklı formattaki maddelerin benzer yapıları ölçtüğü sonucuna ulaşmıştır. Ackerman ve Smith (1988) üst bilişsel düzeyde yer alan davranışların ölçülmesinde yapılandırılmış, alt bilişsel düzeyde yer alan davranışların ölçülmesinde ise çoktan seçmeli maddelerin daha uygun olduğunu belirtmişlerdir. Birgili (2014) yapılandırılmış cevaplı maddelerin bilişsel strateji ve öz kontrol becerilerini daha fazla geliştirdiğini, öğrencilerin yapılandırılmış maddelerde daha fazla çaba, çoktan seçmeli maddelerde ise daha fazla endişe hissettiklerini saptamıştır. Thawabieh (2016) çoktan seçmeli maddelerin yapılandırılmış cevaplı madde türüne göre daha kolay fakat daha az ayırıcı olduğunu, yapılandırılmış cevaplı maddelerden hazırlanan testin ise çoktan seçmeli maddelerle hazırlanan testten daha güvenilir olduğunu belirtmiştir. Demans (1998) madde formatı ve başarı arasında fen alanında düşük bir etkileşim olduğunu, matematik alanında ise etkileşim olmadığını belirtmiştir. Ayrıca yüksek yetenek düzeyinde kız öğrencilerin yapılandırılmış maddelerde, erkek öğrencilerin ise çoktan seçmeli maddelerde daha başarılı olduğunu saptamıştır. Fulcher ve Davidson (2007) öğrenci performanslarının yapılandırılmış (kısa cevaplı) maddelere göre çoktan seçmelilerde daha iyi olduğunu, Beller ve Gafni (2000), Phipps ve Brackbill (2009) ise madde formatının madde parametreleri üzerinde anlamlı etkileri olduğunu ifade etmişlerdir. Ayrıca bu çalışmalarla beraber madde format etkisinin cinsiyet, sosyoekonomik düzey, kültürel özellikler gibi öğrenci özelliklerine göre değişkenlik gösterdiği saptanmıştır (Beller ve Gafni, 2000; Chiu, 2007; Zhang ve Manon, 2000).

Madde özelliklerinin yetenek veya madde parametreleri üzerindeki etkisinin incelendiği çalışmaların büyük bir kısmında iki aşamalı analiz yöntemlerinin kullanıldığı görülmüştür (İlhan vd., 2020; Pepple, Young ve Carroll, 2010; Thawabieh, 2016). Başka bir ifadeyle, bu çalışmalarda öncelikle madde özellikleri veya bireylere ait başarı düzeyleri kestirilmiştir, daha sonra bu parametrelerde ortaya çıkan farklılaşmanın sebepleri madde özellikleri ile ilişkilendirilmiştir. Ayrıca bu çalışmaların çoğunda farklı madde özellikleri arasındaki etkileşiminin incelenmediği, genellikle tek bir madde özelliğine odaklanıldığı görülmüştür. Bu çalışmanın amacı ise veri setinin yapısına daha uygun olan çok seviyeli bir analiz yöntemi olan Açıklayıcı Madde Tepki Kuramı (MTK) modelleri ile madde parametreleri kestirilirken eş zamanlı olarak madde özelliklerinin ve bu özelliklere ait etkileşimin madde güçlüğü

üzerindeki etkisini incelemektir. Ayrıca bu analiz yöntemi sayesinde madde özelliklerine ait etkinin seviyeler arasında rastgele etkisinin ele alınması, başka bir ifadeyle her bireyin aynı düzeyde mi yoksa farklı düzeylerde mi madde özelliklerinden etkilendiği saptanması hedeflenmiştir.

Alanyazında Açıklayıcı MTK modelleri kullanılarak madde özelliklerinin madde parametreleri üzerindeki etkisinin incelendiği çalışmalar mevcuttur. Atar (2011), Açıklayıcı MTK modellerini kullanarak yürüttüğü çalışmada madde konu alanı ve bilişsel alan özelliklerinin madde güçlüğü üzerinde anlamlı etkisi olduğunu saptanmıştır. Hohensinn ve Kubinger (2011), madde formatının (açık uçlu ve çoktan seçmeli) madde özelliklerine etkisini Açıklayıcı MTK modellerinden Rasch model ve Lineer Lojistik Test modeli kullanarak incelemiş, madde formatının madde güçlüğü üzerinde belirgin bir etkisi olduğunu ortaya koymuşlardır. Yine Chiu (2016), Açıklayıcı MTK modellerinden Lineer Lojistik Test modelini kullanarak yürüttüğü çalışmada madde formatının madde güçlüğü üzerindeki etkisinin anlamlı olduğunu belirtmiştir.

Bu bağlamda araştırmanın amacı, madde formatı ve madde bilişsel alan düzeyinin madde güçlük parametresi üzerindeki etkisini Açıklayıcı MTK modelleriyle incelenmektir. Bu amaç doğrultusunda aşağıdaki problemlere yanıt aranmıştır.

1. Madde formatının ve bilişsel alan düzeyinin okuma maddelerinin güçlüğü üzerindeki etkisi bireyler arasında farklılaşmakta mıdır?
2. Çoktan seçmeli ve yapılandırılmış cevaplı okuma maddelerinin güçlükleri arasında anlamlı bir farklılaşma var mıdır?
3. Farklı bilişsel alanlarda yer alan okuma maddelerin güçlükleri arasında anlamlı bir farklılaşma var mıdır?
4. Madde formatı ve madde bilişsel alan düzeyine ait etkileşimin okuma maddelerinin güçlüğü üzerindeki etkisi nasıldır?

Yöntem

Araştırmanın Modeli

Bu çalışmanın amacı madde özelliklerinin güçlük parametresi üzerindeki etkisini incelemek, var olan bir durum belirlemeye çalışmaktır. Bu yüzden bu araştırma betimsel bir çalışmadır (Coe, Waring, Hedges ve Day Ashley, 2021). Betimsel çalışmalarda var olan bir durum tam ve detaylı şekilde ortaya koyulmaya çalışılır.

Araştırma grubu

Araştırmada OECD tarafından üç yılda bir gerçekleştirilen ve her döngüde ağırlıklı olarak farklı bir alanı değerlendiren PISA verileri kullanılmıştır. PISA uygulamasında ağırlıklı alana ait maddeler tüm kitapçıklarda yer almakta ve araştırmaya katılan tüm öğrenciler tarafından cevaplanmaktadır. Ağırlıklı alan dışındaki diğer alanlarda yer alan maddeler ise bazı kitapçıklarda yer almakta ve örneklemden rastgele seçilen öğrenciler tarafından yanıtlanmaktadır. PISA 2015'te ağırlıklı alan fen okuryazarlığı iken diğer alanlar okuma, matematik ve iş birliği öğrenmedir. Dolayısıyla PISA 2015'te okuma ve matematik alanında yer alan maddelere örneklemde yer alan tüm öğrenciler cevap vermemiştir. PISA 2015 Türkiye uygulamasına 5895 (%50,2 kız, %49,5 erkek) öğrenci katılmıştır. Araştırma kapsamında okuma becerilerini ölçen maddeler ele alındığından bu araştırmanın çalışma grubunu PISA 2015 Türkiye uygulamasına katılan ve okuma alanında yer alan maddelere cevap veren 2418 (%50,5 kız, %49,5 erkek) öğrenci oluşturmaktadır.

Veri Toplama Araçları

PISA 2015 Türkiye uygulaması bilgisayar tabanlı yürütülmüş olup toplam 66 kitapçık kullanılmıştır. Bu kitapçıkların 36'sında okuma becerilerine ait maddeler bulunmaktadır. Araştırmaya okuma maddelerini içeren tüm kitapçıklar dâhil edilmiştir. PISA 2015 uygulamasında okuma becerilerini ölçmek için toplam 103 madde mevcuttur. Bu maddelerin 15'i diğer maddelere göre daha kolay olup özel gereksinimi olan öğrenciler için hazırlanmıştır. PISA 2015 Türkiye uygulamasına özel gereksinimi olan öğrenciler dâhil edilmediğinden bu maddelere yer verilmemiştir. Geriye kalan 88 maddenin 81'i ikili, 7'si ise kısmi puanlanmaktadır. Verilerin analizinde kullanılan Açıklayıcı MTK

modelleri ikili puanlanan verilerin analizine uygun olduğundan çalışmaya ikili puanlanan maddeler dâhil edilmiş, üç kategorili olarak puanlanan maddeler analiz dışı bırakılmıştır. Dolayısıyla çalışma 81 okuma maddesi üzerinden yürütülmüştür. Araştırmada kayıp veriler ele alınırken PISA prosedürleri izlenmiş olup öğrencilerin atladığı maddeler yanlış (0), geçersiz veya ulaşılamayan veriler ise kayıp olarak ele alınmıştır.

PISA bilişsel alan testlerinde beş farklı cevap şeklinin kullanıldığı, bu cevap modlarının ise basit çoktan seçmeli (çeldiriciler arasından doğru cevabın seçildiği), kompleks çoktan seçmeli (evet/hayır, doğru/yanlış maddeleri) ve açık uçlu (cevabın öğrenci tarafından yazıldığı) olmak üzere 3 başlık altında ele alındığı görülmüştür (OECD, 2017a). Çalışmada ele alınan 81 maddenin 30'u kompleks çoktan seçmeli, 11'i basit çoktan seçmeli ve 40'ı açık uçlu madde formatındadır. PISA 2015 okuma teknik raporu ve literatür incelendiğinde basit ve kompleks çoktan seçmeli maddelerin çoktan seçmeli, cevabın öğrenciler tarafından oluşturulduğu maddelerin ise yapılandırılmış madde formatında ele alındığı görülmüştür (OECD, 2017b). Bu sebeple bu çalışmada da aynı yol izlenmiş, basit ve kompleks çoktan seçmeli maddeler çoktan seçmeli (41), açık uçlu maddeler (40) ise yapılandırılmış cevaplı maddeler olarak ele alınmıştır. Madde format değişkeni oluşturulurken çoktan seçmeli maddeler 0, yapılandırılmış maddeler ise 1 olarak kodlanmıştır.

Okuma alanında yer alan maddelere ait bilişsel alanlar bilgiye ulaşma ve hatırlama (access ve retrieve), anlama ve yorumlama (integrate and interpret), yansıtma ve değerlendirme (reflect and evaluate) olmak üzere 3 başlık altında ele alınmıştır. Bilgiye ulaşma ve hatırlama basamağında metinden gerekli bilgileri seçme süreci "bilgiye ulaşma", bellekte var olan bilgileri geri getirme zihinsel süreci "hatırlama" olarak ele alınmıştır. Anlama ve yorumlama basamağında metnin içerisindeki neden sonuç ilişkilerini, karşılaştırmaları ve benzerlikleri bulmak "anlama", metnin tamamının veya alt bölümlerinin altında yatan varsayımları ve çıkarımları ayrıntılı bir şekilde açıklamak "yorumlama" basamağını oluşturur. Yansıtma ve değerlendirme basamağında ise öğrencilerden önceki bilgileri ile metinde yer alan bilgileri ilişkilendirmeleri "yansıtma", ortaya koydukları çıkarımları savunmaları, metnin yapısını anlayıp değerlendirmeleri ise "değerlendirme" olarak ele alınmaktadır. Bilişsel alan tanımlamalarında yer alan davranışlar incelendiğinde basamaklar arasında bir hiyerarşinin söz konusu olduğu, bilgiye ulaşma ve hatırlama alanında hazırlanan maddelerin alt bilişsel alanı, yansıtma ve değerlendirme alanında hazırlanan maddelerin ise üst bilişsel alanlı temsil ettiği söylenebilir (OECD, 2017b). PISA 2015 okuma becerilerini ölçmek için hazırlanmış maddelerin bilişsel alanları incelendiğinde bilgiye ulaşma ve hatırlama basamağında 20, anlama ve yorumlama basamağında 45, yansıtma ve değerlendirme basamağında ise 16 madde olduğu görülmüştür. Madde bilişsel alanına ait değişken oluşturulurken değerlendirme basamağı referans olarak alınmış, bilgiye ulaşma ve hatırlama (bilgi) ve anlama ve yorumlama (anlama) olmak üzere iki değişken oluşturulmuştur. Bu değişkenlerin kodlanmasında "bilgi" değişkeni oluşturulurken bilişsel alanı bilgiye ulaşma ve hatırlama olan maddeler 1, anlama ve değerlendirme olan maddeler 0; "anlama" değişkeni oluşturulurken bilişsel alanı anlama olan maddeler 1, bilgi ve değerlendirme olan maddeler 0 olarak kodlanmıştır.

Verilerin Toplanması

Araştırmanın verileri PISA 2015 okuma becerileri bilişsel alt testine ait olup <https://www.oecd.org/pisa/data/> adresinden indirilmiştir.

Verilerin Analizi

Özellikle uluslararası uygulanan geniş ölçekli sınav analizleri MTK modelleri ile yürütülmekte, madde parametreleri ve bireylere ait örtük özellikler kestirilmektedir (Lord, 1980). Bu modeller yardımıyla bireylerin performansları yetenek, madde özelliklerini ise kullanılan modele göre güçlük ve ayırıcılık cinsinden tanımladığı için "Tanımlayıcı MTK Modelleri" olarak adlandırılırlar (De Boeck ve Wilson, 2004). Tanımlayıcı MTK modellerinde birey ve madde parametrelerinde ortaya çıkan farklılıklar ele alınmaz. Bu modeller genelleştirilmiş doğrusal/doğrusal olmayan karma modeller çerçevesinde ele alınarak birey veya madde parametrelerinde ortaya çıkan farklılaşma birey, madde veya her ikisinin etkileşimine ait değişkenlerin modele dâhil edilmesiyle eş zamanlı olarak açıklanabilir (De Boeck ve Wilson, 2004). Madde ve birey parametreleri kestirilirken aynı zamanda bu parametreler arasında

ortaya çıkan farklılaşmanın ele alındığı modeller “Açıklayıcı MTK Modelleri” olarak adlandırılır (De Boeck ve Wilson, 2004). Açıklayıcı MTK modelinde öğrencilere ait yanıtlar tekrarlı ölçümler olarak ele alınabilir. Başka bir ifade ile bir öğrencinin bir testte yer alan maddelere verdiği yanıtlar öğrencinin içerisinde kümelenmiş kabul edilir. Böylelikle veri seti çok seviyeli olarak ele alınarak (1. seviye madde yanıtları, 2. seviye öğrenciler olarak) yordayıcı değişkenlere ait etkinin seviyeleri farklılaşıp farklılaşmadığı incelenebilir.

Genelleştirilmiş doğrusal/doğrusal olmayan karma modellerin özel durumları olan Açıklayıcı MTK modellerinde sık kullanılan dört farklı model mevcuttur. İlk model çifte tanımlayıcı model olarak da bilinen Rasch modelidir. Bu modelde madde güçlükleri ve yetenek parametresi kestirilirken bu parametreler arasında ortaya çıkan farklılaşmayı yordayan herhangi bir değişken yer almaz. İkinci model madde güçlükleri arasında ortaya çıkan farklılaşmayı ele almak için madde özelliklerinin yordayıcı değişken olarak dâhil edildiği fakat birey düzeyinde herhangi bir yordayıcı değişken içermeyen Lineer Lojistik Test Modelidir. Üçüncü model birey parametrelerinde ortaya çıkan farklılaşmayı ele almak için birey özelliklerinin dâhil edildiği fakat madde düzeyinde herhangi bir değişken içermeyen Örtük Regresyon Rasch Modelidir. Çifte açıklayıcı model olarak bilinen son model ise madde ve birey parametreleri arasında ortaya çıkan farklılaşmayı ele almak için hem madde hem de birey düzeyinde yordayıcı değişkenler içeren Örtük Regresyon Lineer Lojistik Test Modelidir.

Araştırma kapsamında madde formatının ve madde bilişsel alanının madde güçlüğü üzerindeki etkisinin incelenmesi, madde güçlük parametresinde ortaya çıkan farklılaşmanın ele alınması amaçlanmaktadır. Bu yüzden çalışmanın analizlerinde madde tanımlayıcı model olarak da bilinen Lineer Lojistik Test Modeli (LLTM) kullanılmıştır. Bu modelde madde güçlük parametresi bağımlı değişken olarak ele alınır ve bağımlı değişken üzerinde etkili olduğu düşünülen bağımsız değişkenlerle aradaki ilişki regresyon eşitliği ile kurulur. Ayrıca LLTM’de artık terimler (residual term) de ele alınabilir. Başka bir ifade ile madde güçlüklerine ait etkinin seviyeler arasında farklılaşmasına izin verilerek daha esnek kestirimler yapılabilir (De Boeck, 2008).

Verilerin analizine madde ve birey parametrelerini rastgele kestiren ve hiçbir yordayıcı değişken içermeyen boş model (M0) ile başlanılmıştır. Daha sonra madde güçlükleri arasındaki farklılığı açıklamak üzere madde formatı değişkenine ait sabit etki modele (M1) dâhil edilmiştir. Madde formatına ait etkinin bireyler arasında farklılaşıp farklılaşmadığını belirlemek için bu değişkene ait rastgele etki (M2) modellenmiştir. Aynı işlem maddelere ait bilişsel alan değişkeni için de yapılmıştır (M3-M4). Daha sonra kurulan modellere ait model uyum indeksleri karşılaştırılmış, madde formatı ve madde bilişsel özelliklerinin bireyler arasında farklılaşıp farklılaşmadığına karar verilmiştir. Son olarak da tüm değişkenler modele dâhil edilerek M5 modeli kurulmuş, madde formatı ve madde bilişsel alanının madde güçlüğü üzerindeki etkileri incelenmiştir. Analizler R programında lme4 (Bates, Maechler, Bokler ve Walker, 2018) paketi aracılığıyla yürütülmüş, Maximum Likelihood (Laplace Approximation) kestirim yöntemi kullanılmıştır.

Geçerlik ve Güvenirlik

Kurulan modellere ait model uyum indeksleri bulgular bölümünde verilmiştir.

Araştırma ve Yayın Etiği

Bu çalışmada “Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi” kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan “Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler” başlığı altında belirtilen eylemlerden hiçbiri gerçekleştirilmemiştir.

Etik Kurul İzni

Çalışmada kullanılan PISA 2015 veri seti erişime açık olduğundan etik kurul izni gerekmemektedir.

Bulgular

Analizlere madde ve birey düzeyinde hiçbir yordayıcı değişken içermeyen boş model (M0) ile başlanmıştır. Bu model ile maddelere ait güçlük ve bireylere ait yetenek parametreleri kestirilmiş, sırasıyla bu kestirime ait varyans değerlerinin 0,88 ve 1,55 lojit olduğu görülmüştür. Maddelere ait güçlük değerleri lojit ölçeğinde -3,32 ve 4,30 aralığında değişmekte olup bu değerlerin ortalaması 0,16'dır.

Daha sonra madde formatının sabit (M1) ve rastgele etkisini (M2) içeren modeller ile madde bilişsel alan değişkeninin sabit (M3) ve rastgele etkisini içeren (M4) modeller kurulmuştur. M1 modeli M2 modelinin, M3 modeli de M4 modelinin içinde kümelendiğinden model uyum indekslerini karşılaştırmak için ki-kare fark testi uygulanmış, aynı zamanda AIC ve BIC değerleri incelenmiştir. Tablo 1'de bu modellere ait model uyum indeksleri sunulmuştur.

Tablo 1.

Model Uyum İndeksleri

Alan	Model	AIC	BIC	logLik	χ^2
Format	M1	71983	72001	-35927	
	M2	71862	71898	-35927	125.11***
Bilişsel alan	M3	71472	71499	-35733	
	M4	71420	71492	-35702	61.78 ***

* p<.05; ** p<.01; *** p<.001

Tablo 1 incelendiğinde M2 modelinin M1, M4 modelinin de M3 modeline göre daha düşük AIC ve BIC değerlerine sahip olduğu görülmüştür. Ayrıca her iki değişken için kurulan sabit ve rastgele modeller arasında yapılan ki kare fark testi sonucu madde formatı ($\chi^2 = 125,11$, $p < ,000$) ve bilişsel alan ($\chi^2 = 61,78$, $p < ,000$) değişkeninin bireyler arasında rastgele etkisini ele alan modellerin daha iyi uyum sağladığı görülmüştür. Bu sonuç madde özelliklerine ait etkinin bireyler arasında sabit olmadığı, başka bir ifade ile öğrencilerin madde formatı ve madde bilişsel alanından farklı düzeylerde etkilendiği şeklinde yorumlanmıştır.

Analizlere madde formatı ve bilişsel alan düzeyinin madde güçlüğü üzerindeki etkisini incelemek amacıyla bu değişkenlerin ve bu değişkenlere ait kesişimin dâhil edildiği M5 modeli ile devam edilmiştir. M5 modeli ile madde güçlük parametresine ait varyans 1,02'ye düşmüştür. Bu düşüş modele dahil edilen değişkenlerle madde güçlüğündeki farklılıkların %34 oranında açıklandığı anlamına gelmektedir. Model 5'e ait sabit etkiler Tablo 2'de sunulmuştur.

Tablo 2.

Madde Formatı ve Bilişsel Alan Özelliklerinin Madde Güçlük İndeksine Etkisi

Alan	Tahmin	S.H	Z değeri
Format	0,08***	0,02	-5,08
Bilgi	-1,11***	0,07	-16,12
Anlama	0,21***	0,01	11,17
Format-Bilgi	0,43***	0,04	10,16
Format-Anlama	-0,04	0,03	1,32

* p < ,05; ** p < ,01; *** p < ,001

Tablo 2 incelendiğinde, çoktan seçmeli maddeler referans olarak alındığından, açık uçlu maddelerin çoktan seçmeli maddelere göre 0,08 lojit daha zor olduğu ve bu farklılığın anlamlı olduğu görülmüştür. 0,08 değerinin üsseli alındığında ($\exp(0,08)$) odds oranı elde edilir. Odds oranı bir maddenin doğru cevaplanma ihtimalinin yanlış cevaplanma ihtimaline oranıdır. Maddelerin doğru yanıtlanma olasılığı hesaplandığında, çoktan seçmeli bir maddenin doğru cevaplanma ihtimali %50 olduğunda açık uçlu bir maddenin doğru cevaplanma olasılığının yaklaşık %48 olduğu söylenebilir.

Bilişsel alan değişkenleri incelendiğinde, değerlendirme maddeleri referans olarak alındığından, bilişsel alanı bilgi olan maddelerin değerlendirme olanlara göre 1,11 lojit daha kolay olduğu ve bu farklılığın anlamlı olduğu saptanmıştır. Doğru cevap verme olasılıkları incelendiğinde ise bilişsel alanı değerlendirme olan maddelerin doğru cevaplanma olasılığı %50 olduğunda bilişsel alanı bilgi olan maddeler için bu olasılık yaklaşık %75'e yükselecektir. Bilişsel alanı anlama olan maddelerin bilişsel alanı değerlendirme olan maddelerden 0,20 lojit daha zor olduğu ve bu farklılığın anlamlı olduğu saptanmıştır. Doğru cevaplanma olasılıkları incelendiğinde ise bilişsel alanı değerlendirme olan maddelerin doğru cevaplanma olasılığı %50 olduğunda bilişsel alanı anlama olan maddeler için bu olasılığın yaklaşık %45'e düşeceği söylenebilir.

Çalışmanın bir başka çıktısı da madde formatı ve madde bilişsel alanının kesişimleri ile ilgilidir. Çalışmada en kolay maddelerin bilişsel alanı bilgi olan maddeler olmasına rağmen format-bilgi kesişimi incelendiğinde bir maddenin bilişsel alanı bilgi ve madde formatı açık uçlu olduğunda 0,443 lojit daha zor olacağı, hatta en zor maddelerin bilişsel alanı bilgi olan açık uçlu maddeler olduğu saptanmıştır.

Bilişsel alanı anlama olan maddelerin diğer alanlara göre daha zor olduğu, format-anlama kesişimi incelendiğinde ise bu maddelerin açık uçlu sorulmasının maddeyi 0,04 lojit daha kolay hale getireceği görülmüştür. Her ne kadar bu etki anlamlı olmasa da bilişsel alanı anlama olan maddelerinin anlamlı olarak daha zor olmasına rağmen bu maddelerin açık uçlu sorulmasıyla ortalama zorluğun düşeceği hatta miktar kolaylaşacağı görülmektedir.

Tartışma ve Sonuç

Bu çalışmada madde özelliklerinin madde güçlüğü üzerindeki etkisi Açıklayıcı MTK modeli ile incelenmiştir. Elde edilen sonuçlar özetlenecek olursa format değişkeninin madde güçlüğü üzerinde anlamlı etkisi olduğu, açık uçlu maddelerin çoktan seçmeli maddelere göre daha zor olduğu görülmüştür. Bu bulgu literatürde yer alan çalışmaların sonuçlarıyla paralellik göstermektedir (Fuchler ve Davidson, 2007; Griffo, 2011; İlhan ve diğerleri, 2020). Yapılandırılmış cevaplı maddelerde öğrencilerden cevabı tasarlayıp yazmalarının istenmesi, öğrencilerin bu formata alışık olmamaları veya öğrencilerin yazma becerilerinin iyi olmaması bu sonucu doğurmuş olabilir. Bir başka sebep de sayısal alanlara göre sözel alanlarda madde formatı etkisinin daha belirgin olması olabilir (Traub ve Fisher, 1997). Literatürde çoktan seçmeli ve yapılandırılmış maddelerden oluşan testlerin benzerlik ve farklılıklarıyla ilgili çeşitli tartışmalar bulunmaktadır. Ackerman ve Smith (1988) ve Wainer ve Thissen (1993) yapılandırılmış ve çoktan seçmeli maddelerin temelde benzer yapıları ölçtüğünü, Walstad ve Becker (1994) ise ekonomik, zaman ve maliyet açısından daha kullanışlı olan çoktan seçmeli maddelerin tercih edilmesi gerektiğini belirtmiştir. Yine Bacon (2003) üst düzey bilişsel özellikleri test etmeyi sevse bile çoktan seçmeli maddelerin daha uygun bir ölçüm aracı olduğunu ifade etmiştir. Buna karşılık Becker ve Johnston (1999), Anderson ve arkadaşları (2000), Dufresene, Leonard ve Gerace (2002) çoktan seçmeli ve yapılandırılmış cevaplı maddelerin farklı bilişsel boyutları ölçtüğünü, Martinez (1999) ve Hancock (1994) ise bu görüşün kısmen doğru olduğunu, alt bilişsel düzeyleri ölçen madde formatlarının benzer olduğunu fakat üst bilişsel düzeylerin ölçülmesinde genel bir eşitliğin olmadığını belirtmişlerdir.

Çalışmada bilgi basamağında yer alan maddelerin en kolay, anlama basamağında yer alan maddelerin ise en zor madde türü olduğu görülmüştür. PISA'da kullanılan maddelere ait bilişsel alan davranışları incelendiğinde bilgi basamağının alt, değerlendirme basamağının ise üst bilişsel becerileri temsil ettiği söylenebilir. Bazı görüşlere göre üst bilişsel düzeyde yer alan bir davranışın gerçekleşmesi için alt biliş düzeyinde yer alan davranışların kazanılması gerektiği (Bloom, Krathwohl ve Masia, 1956), bazı görüşlere göre ise böyle bir hiyerarşinin söz konusu olmadığı, alt düzeyde yer alan bir davranış tamamlanmadan da üst düzeyde yer alan bir davranışın kazanılabileceği ifade edilmiştir (Ormel, 1979). Çalışmada anlama basamağında yer alan maddelerin değerlendirme basamağından daha zor algılanması, bilişsel alanlar arasında net bir hiyerarşinin olmadığı görüşünü desteklemektedir. Ayrıca bu sonucun ortaya çıkması madde içerikleriyle veya bu alanda yer alan kazanımlarla (karşılaştırma, ilişkilendirme, çıkarım yapma gibi) ilişkili olabilir.

Çalışmada elde edilen bir başka sonuçta madde formatı ve madde bilişsel özelliklerinin kesişimleri ile ilgilidir. Elde edilen sonuçlar incelendiğinde en kolay madde türünün bilgi basamağında

yer alan maddeler olmasına rağmen bu maddelerin açık uçlu sorulmasının maddelerin güçlüklerini arttırdığı, hatta madde güçlüğü en fazla arttıran değişken olduğu saptanmıştır. Benzer sonuç anlama-format değişkeni için de elde edilmiştir. Anlama alanında yer alan maddelerin diğer basamakta yer alan maddelere göre daha zor olmasına rağmen bu maddelerin açık uçlu sorulmasının madde güçlüğü düşürdüğü, başka bir anlatımla bir miktar kolaylaştırdığı görülmüştür. Bu sonuç açık uçlu maddelerin üst bilişsel, çoktan seçmeli maddelerin ise alt bilişsel alanlarda yer alan kazanımları ölçmede daha iyi olduğu görüşünü desteklemektedir (Ackerman ve Smith, 1988; Melovitz Vasan, DeFouw, Holland ve Vasan, 2018).

Bu çalışmada okuma becerileri alanında yer alan maddelerin madde özellikleriyle ilişkisi incelenmiştir. Bundan sonra yapılacak olan çalışmalar da farklı alanlar (sayısal ve sözel) ele alınarak madde formatının etkisi karşılaştırılabilir. PISA verilerinin analizinde madde güçlükleriyle beraber madde ayırt edicilikleri de MTK modelleri ile incelenmektedir. Bu çalışmada madde özelliklerinin sadece madde güçlüğü üzerindeki etkileri incelenmiştir. Bundan sonra yapılacak olan çalışmalarda madde özelliklerinin madde güçlüğü yanında madde ayırt ediciliklerini nasıl etkilediği araştırılabilir. Ayrıca araştırmada kompleks ve basit çoktan seçmeli maddeler çoktan seçmeli maddeler başlığı altında ele alınmıştır. Gelecek çalışmalarda basit, kompleks ve açık uçlu madde formatlarının madde parametreleri üzerindeki etkisi ayrı ayrı incelenip karşılaştırılabilir.

Araştırma ve Yayın Etiği

Bu çalışmada “Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi” kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan “Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler” başlığı altında belirtilen eylemlerden hiçbiri gerçekleştirilmemiştir.

Etik Kurul İzni

Çalışmada kullanılan PISA 2015 veri seti erişime açık olduğundan etik kurul izni gerekmemektedir.

Yazarların Katkı Oranı

Yazarlar çalışmaya eşit derecede katkı sağlamışlardır.

Çıkar Çatışması

Bu çalışmada çıkar çatışması teşkil edebilecek herhangi bir durum söz konusu değildir.

Kaynaklar

- Ackerman, T. A., & Smith, P. L. (1988). A comparison of the information provided by essay, multiple-choice, and free-response writing tests. *Applied Psychological Measurement, 12*(2), 117-128.
- Anderson, L., Krathwohl, D., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., et al. (2000). *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives*. Abridged Edition: Allyn & Bacon.
- Bacon, D. R. (2003). Assessing learning outcomes: A comparison of multiple-choice and short-answer questions in a marketing context. *Journal of Marketing Education, 25*(1), 31-36.
- Badger, E., & Thomas, B. (1991). Open-ended questions in reading. *Practical Assessment, Research, and Evaluation, 3*(1), 4.
- Bates, D., Maechler, M., Bokler, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.
- Becker, W. E., & Johnston, C. (1999). The relationship between multiple choice and essay response questions in assessing economics understanding. *Economic Record, 75*(4), 348-357.
- Beller, M., & Gafni, N. (2000). Can item format (multiple choice vs. open-ended) account for gender differences in mathematics achievement? *Sex Roles, 42*(1), 1-21.
- Bennett, R. E., Ward, W. C., Rock, D. A., & LaHart, C. (1990). *Toward a framework for constructed-response items*.

- Bennett, R. E., Rock, D. A., Braun, H. I., Frye, D., Spohrer, J. C., & Soloway, E. (1990). The relationship of expert-system scored constrained free-response items to multiple-choice and open-ended items. *Applied Psychological Measurement, 14*(2), 151-162.
- Ben-Simon, A., Budescu, D. V., & Nevo, B. (1997). A comparative study of measures of partial knowledge in multiple-choice tests. *Applied Psychological Measurement, 21*(1), 65-88.
- Bible, L., Simkin, M. G., & Kuechler, W. L. (2008). Using multiple-choice tests to evaluate students' understanding of accounting. *Accounting Education: An International Journal, 17*(S1), S55-S68.
- Birgili, B. (2014). *Open ended questions as an alternative to multiple choice: Dilemma in Turkish examination system* (Master's thesis). Middle East Technical University Institute of Social Sciences, Ankara.
- Bloom BS, Krathwohl DR, & Masia BB (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York: McKay.
- Brown, G. A., Bull, J., & Pendlebury, M. (2013). *Assessing student learning in higher education*. Routledge.
- Bush, M. (2001). A multiple choice test that rewards partial knowledge. *Journal of Further and Higher Education, 25*(2), 157-163.
- Coe, R., Waring, M., Hedges, L., & Day Ashley, L. (Eds.) (2021). *Research methods and methodologies in education*. SAGE Publications. <https://us.sagepub.com/en-us/nam/research-methods-and-methodologies-in-education/book271175#description>
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Toronto: Holt, RineHart, and Winston Inc.
- Cruickshank, D. L., Bainer, D. L., & Metcalf, K. K (1995). *The act of teaching*. New York: McGraw-Hill.
- De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: a generalized linear and nonlinear approach*. New York, NY: Springer Press.
- De Boeck, P. (2008). Random item IRT models. *Psychometrika, 73*(4), 533-559.
- Demir, E. (2010). *Uluslararası öğrenci değerlendirme programı (PISA) bilişsel alan testlerinde yer alan soru tiplerine göre Türkiye’de öğrenci başarıları* (Yayımlanmamış yüksek lisans tezi). Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Dufresne, R. J., Leonard, W. J., & Gerace, W. J. (2002). Marking sense of students' answers to multiple-choice questions. *The Physics Teacher, 40*(3), 174-180.
- Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37*(6), 359-374.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. London and New York: Routledge.
- Gardner, R.C., Tremblay, P.F. & Masgoret, A.M. (1997). Towards a Full Model of Second Language Learning: An Empirical Investigation. *The Modern Language Journal, 81*(3), 344-362.
- Geer, J. G. (1988). What do open-ended questions measure? *Public Opinion Quarterly, 52*(3), 365-367.
- Hancock, G. R. (1994). Cognitive complexity and the comparability of multiple-choice and constructed-response test formats. *The Journal of Experimental Education, 62*(2), 143-157.
- Haynie, W. (1994). Effect of Multiple – Choice & short answer test on delayed retention learning. *Journal of Technology Education, 6*(1). 32-44
- Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*(3), 235-266.
- Hurd, A. W. (1932). Comparisons of short answer and multiple choice tests covering identical subject content. *The Journal of Educational Research, 26*(1), 28-30.
- Jennings, S., & Bush, M. (2006). A comparison of conventional and liberal (free-choice) multiple-choice tests. *Practical Assessment, Research, and Evaluation, 11*(1), 8.
- Kufahi, T.(2003). *Measurement & evaluation in special education*. Amman: Dar Almasira.
- Lee, H.-S., Liu, O. L. ve Linn, M. C. (2011). Validating measurement of knowledge integration in science using multiple-choice and explanation items. *Applied Measurement in Education, 24*(2), 115-136.

- Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Lukhele, R., Thissen, D., & Wainer, H. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement, 31*(3), 234–250.
- Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist, 34*(4), 207-218.
- Melovitz Vasan, C. A., DeFouw, D. O., Holland, B. K., & Vasan, N. S. (2018). Analysis of testing with multiple choice versus open-ended questions: Outcome-based observations in an anatomy course. *Anatomical Sciences Education, 11*(3), 254-261.
- Organisation for Economic Co-operation and Development [OECD]. (2017a). *PISA 2015 technical report*. Paris, France: OECD. Retrieved from <https://www.oecd.org/pisa/data/2015-technical-report/>.
- Organisation for Economic Co-operation and development [OECD]. (2017b). *PISA 2015 technical report*. Paris, France: OECD. Retrieved from <https://www.oecd.org/pisa/data/2015-technical-report/>.
- Ormell, C. P. (1974). Bloom's taxonomy and the objectives of education. *Educational Research, 17*, 3-18.
- Osterlind, S. J. (1998). *Constructing test items: Multiple-choice, constructed-response, performance, and other formats*. Dordrecht, Netherlands: Kluwer Academic.
- Robbins, A. (1995). *İçindeki devi uyandır*. (Çev. B. Çorakçı Dişbudak). İstanbul: İnkılap Yayınevi.
- Ruch, G. M., & Stoddard, G. D. (1925). Comparative reliabilities of five types of objective examinations. *Journal of Educational Psychology, 16*(2), 89.
- Pepple, D. J., Young, L. E., & Carroll, R. G. (2010). A comparison of student performance in multiple-choice and long essay questions in the MBBS stage I physiology examination at the University of the West Indies (Mona Campus). *American Journal of Physiology - Advances in Physiology Education, 34*(2), 86–89.
- Phipps, S. D., & Brackbill, M. L. (2009). Relationship between assessment item format and item performance characteristics. *American Journal of Pharmaceutical Education, 73*(8).
- Pollack, J. M., Rock, D. A., & Jenkins, F. (1992). *Advantages and disadvantages of constructed-response item formats in large-scale surveys*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.
- Powell, J. L., & Gillespie, C. (1990). *Assessment: all tests are not created equally*. Paper presented at the Annual Meeting of the American Reading Forum, Sarasota.
- Traub, R. E., & Fisher, C. W. (1997). On the equivalence of constructed-response and multiple-choice tests. *Applied Psychological Measurement, 1*(3), 355-369.
- Van den Bergh, H. (1990). On the construct validity of multiple-choice items for reading comprehension. *Applied Psychological Measurement, 14*(1), 1-12.
- Ventouras, E., Triantis, D., Tsiakas, P., & Stergiopoulos, C. (2010). Comparison of examination methods based on multiple-choice questions and constructed-response questions using personal computers. *Computers & Education, 54*(2), 455-461.
- Wainer, H., & Thissen, D. (1993). Combining multiple-choice and constructed-response test scores: Toward a Marxist theory of test construction. *Applied Measurement in Education, 6*(2), 103-118.
- Walstad, W. B., & Becker, W. E. (1994). Achievement differences on multiple-choice and essay tests in economics. *The American Economic Review, 84*(2), 193–196.
- Walstad, W. B. (1998). Multiple choice tests for the economics course. In W. B. Walstad & P. Saunder (Eds.). *In teaching undergraduate economics: A handbook for instructors* (pp. 287-304), New York: McGraw- Hill.
- Zeidner, M. (1987). Essay versus multiple-choice type classroom exams: the student's perspective. *The Journal of Educational Research, 80*(6), 352-358.

Extended Abstract

Introduction

In educational and psychological tests, the test type, the purpose of use, the measured behavioral characteristics and the type of the item used in the test are the basic features that make up the structure of the test. It is crucial to select the proper item format and consider whether the students will benefit from it.

It is accepted that the responses given to the items in the tests reflect the degree of individuals having the characteristic that is intended to be measured. Therefore, a major step in the test development is the creation of items that can best measure the measured. When deciding on the item format, different variables such as learning outcomes, students' readiness levels and ages, the purpose of the test, the nature of the content/structure to be measured, and the experience of the testers should be considered.

Previous research has investigated how item characteristics affect achievement or item parameters, but it was seen that two-stage analysis methods were used in most of these studies. In these studies, first, item characteristics or individual achievement levels were estimated, and then the reasons for the differentiation in these parameters were associated with the item characteristics. Additionally, most of these studies focused on a single item feature, where the interaction between different item properties was not examined. The aim of the current research is to examine the effect of item properties and the interaction of these properties on item difficulty while estimating item parameters with Explanatory IRT models, which is a multi-level analysis method that is more suitable for the structure of the data set.

Additionally, this analysis method aimed to deal with the random effect of the item properties between the levels, in other words, to determine whether each student is affected by the item properties at the same level or at different levels. In line with this purpose, the following research questions are substantiated in this paper:

1. Do the effects of item format and cognitive level on the difficulty of reading items differ among individuals?
2. Do item formats have an effect on item difficulties?
3. Do the different cognitive levels of items affect the item difficulty parameter?
4. What is the interaction effect of the item format and item cognitive level on the difficulty of reading items?

Method

The research was conducted with 2418 (50.5% female, 49.5% male) Turkish students who participated in the PISA 2015 study. The PISA 2015 reading skills cognitive subtest data is used, which can be downloaded on <https://www.oecd.org/pisa/data/>. The study includes 81 reading items (30 complex multiple-choice items, 11 simple multiple-choice items and 40 open-response items). Simple and complex multiple-choice items are considered multiple choice (41), and open-response items (40) are considered items with constructed response.

The cognitive level of the items in reading literacy are access and retrieve, integrate and interpret, reflect and evaluate. There are 20 items at the access and retrieve (access), 45 items at the integrate and interpret (integrate), and 16 items at the reflect and evaluate (evaluate) level. The evaluate level is taken as a reference, and two variables are defined as access-evaluate (access) and integrate-evaluate (integrate).

Linear Logistic Test Model (LLTM) is used in the analysis of the research. The item difficulty parameter is considered to be the dependent variable in the model, and the relationship between the independent variables that are thought to be effective on the dependent variable is established by regression. Residual terms are also considered in LLTM which is more flexible estimations are made by allowing the effect of item difficulties to differ among levels.

Result and Discussion

The results showed that the format variable has a significant effect on item difficulty, and open-ended items are more difficult than multiple-choice items which is similar to the results of other studies.

Open-response and multiple-choice items basically measure similar properties and multiple-choice items that are more economical should be preferred. Becker & Johnston (1999) stated that items with selected-response and constructed-response measure different cognitive aspects, while Martinez (1999) & Hancock (1994) stated that this view is partially true, with selected-response and constructed-response measuring sub-cognitive aspects. They found out that the items were similar but there was no general equality in the measurement of cognitive levels.

This study found out that the items in the access level are the easiest item type. The items in the integrate level are the most difficult item type. This finding supports the view that there is no clear hierarchy between cognitive domains. Another result is related to the interactions of item format and item cognitive levels. Although the easiest item type is access level items, asking these items with constructed response increased the difficulty of the items. Similar results are obtained for the integrate-format variable. Although the integrate items are more difficult than the other items, it is observed that asking the items in this area with constructed response reduced the item difficulty. This result supports the view that constructed-response items are more suitable for measuring metacognitive domains while selected items are better for measuring achievements in sub-cognitive domains.