

PISA 2012 Okuduğunu Anlama Testine Verilen Yanıtların Boyutluluğunun İki Faktör Modeline Dayalı Olarak İncelenmesi

Seval Kula Kartal¹

Öz

Bu araştırmanın amacı yanıtlayıcıların PISA 2012 okuduğunu anlama testi maddelerine verdiği yanıtlardan elde edilen veri setinin boyutluluğunun iki parametrelili lojistik model ve iki faktör modeliyle elde edilen madde parametreleri ve madde düzeyinde model veri uyumu istatistiklerinin karşılaştırılması yoluyla incelenmesidir. PISA 2012 Türkiye uygulamasında dört okuma metnine dayalı olarak geliştirilmiş 14 madde içeren 12 numaralı test formu 284 kişilik öğrenci grubu üzerinde uygulanmıştır. Madde Tepki Kuramı modelleri arasındaki karşılaştırmalar madde ayırt edicilik parametrelerine, $S-\chi^2$ madde uyum istatistiklerine ve açıklanan ortak varyans değerlerine dayalı olarak yapılmıştır. Analizlerin sonuçları, genel boyut ayırt edicilik parametrelerinin tek boyutlu iki parametrelili lojistik model madde parametreleriyle benzer olduğunu göstermiştir. İki faktör modeli madde düzeyinde model veri uyumunda tek boyutlu modele göre bir miktar iyileşme sağlamış olsa da bu iyileşmenin önemli olmadığı bulunmuştur. Madde model veri uyumu açısından iki model de benzer sonuçlar sağlamıştır. Bu bulgulara dayalı olarak, genel boyuta karşılık gelen okuduğunu anlama becerisinin test verisinin altında yatan baskın boyut olduğu, ortak metinlerden kaynaklı varyansın küçük ve önemsiz olduğu, verinin (yaklaşık) tek boyutluluk varsayımını sağladığı sonucuna varılmıştır.

Anahtar Sözcükler

Boyutluluk
Madde tepki kuramı
İki faktör modeli
Okuduğunu anlama

Makale Hakkında

Gönderim Tarihi

13 Ekim 2018

Kabul Tarihi

23 Ocak 2019

Makale Türü

Araştırma Makalesi

DOI: 10.12984/egcedf.470194

Examining Dimensionality of Responses to the PISA 2012 Reading Comprehension Test Based on the Bifactor Model

Abstract

The objective of this research is to examine dimensionality of the data set obtained from the test takers' responses to the PISA 2012 reading literacy test by comparing item parameters, and item level model data fits estimated based on the two-parameter logistic model and the bifactor model. The PISA 2012 Reading Literacy Test Booklet 12, including fourteen items related to four reading texts, was conducted on a group of 284 students. Model comparisons were done based on item discrimination parameters, $S-\chi^2$ item fit statistics, and the index of explained common variance calculated based on item parameters. Results of the analyses indicate that item discrimination parameters estimated on the general dimension are similar to the two-parameter logistic model item parameters. The bifactor model provided some improvement on the item level fit over the one-dimensional model, however this improvement is not meaningful. Both models produced similar results in terms of the item data fit. Based on these findings, it was concluded that the general dimension representing reading comprehension skill is the dominant dimension underlying the data, and the text effect is small enough to accept that the data holds (essential) the unidimensionality assumption.

Keywords

Dimensionality
Item response theory
Bifactor model
Reading comprehension

Article Info

Received

October 13, 2018

Accepted

January 23, 2019

Article Type

Research Paper

Atf: Kartal, S. K. (2019). PISA 2012 Okuduğunu anlama testine verilen yanıtların boyutluluğunun iki faktör modeline dayalı olarak incelenmesi. *Ege Eğitim Dergisi*, 20(1), 67-77. doi: 10.12984/egcedf.470194

¹ Pamukkale Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Türkiye, sevalk@pau.edu.tr

Extended Abstract

Introduction

Investigating statistical structure and dimensionality of an item response data obtained from the interactions between test takers and test items is important and necessary in the test development process. Therefore, it is possible to find a great deal psychometric studies focusing on the dimensionality concept and the dimensionality assessment methods. In reading comprehension assessments, it is very common to use passages. When passages are utilized to assess students' reading skills, multiple questions that are linked to the same passage are developed. If the questions are nested within the passages, it is possible to have statistical dependences among test taker's items responses, because items are based on the same passage, and test takers' responses to some questions may affect their responses to other questions (So, 2010). Different dimensionality assessment methods may result in different solutions. In case of having inconsistent findings provided by various methods, the bifactor model can help researchers to make more informed decisions on accepting versus rejecting the unidimensionality assumption.

The bifactor model assumes the multidimensionality in the form of a bifactor structure. Bifactor structure allows test items to have loadings on the general dimension and one of the sub-dimensions. The general dimension reflects the targeted trait with the measurement tool and explains the covariance shared by all items. Sub-dimensions explain covariances among items belonging to the same cluster. In addition, it provides a way to investigate the degree of deviation from unidimensionality by comparing item parameters estimated on the general dimension with item parameters estimated by the unidimensional model. The similarity between item parameters indicates that item response data does not largely differ from unidimensionality (Luo, & Al-Harbi, 2016; Reise, Morizot, & Hays, 2007). It is regarded as important to investigate dimensionality of the data obtained from passage-based items, because this format is likely to violate fundamental assumptions of the item response theory such as the local independence and the unidimensionality. Therefore, the objective of this research is to examine dimensionality of the data set obtained from test takers' responses to the PISA 2012 reading comprehension test items by comparing item parameters, item level model data fits and the explained common variance index of the two parameter logistic model and the bifactor model.

Method

The research is a quantitative study aiming to reveal dimensionality of the data that is obtained by applying the PISA 2012 reading comprehension achievement test on students by comparing item parameter estimations and item fits of the unidimensional item response model and the bifactor model. The PISA 2012 Reading Literacy Test Booklet 12 includes fourteen items related to four different reading texts. Booklet 12 was conducted on 284 students. All model parameters and model data fit statistics were estimated using the *mirt package* (Chalmers, 2012) on the R which is an open-source statistical program. Model comparisons were done based on item discrimination parameters, $S-\chi^2$ item fit statistics, the explained common variance index and the omega coefficients calculated based on the unidimensional and the multi-dimensional models.

Findings

In order to reveal the passage effect on item dependences and the strength of reading comprehension skill as the general dimension, item parameter estimations were done based on the bifactor model and the two-parameter logistic item response model. In the bifactor model, test items were allowed to have loadings (or discrimination parameter under IRT) on the general dimension representing the reading comprehension skill and sub-dimensions representing the reading texts. In the two-parameter logistic model, items were forced to have loadings on one dimension. The similarity of item discrimination parameters estimated by the two-parameter logistic model with item discriminations estimated on the general dimension by the bifactor model provides evidence to the existence of (essential) unidimensionality and approximate simple structure in the data. In this study, it was found that the highest difference among the item parameters estimated by the models is 0.45. The similarity among item parameter estimations provided by the models supports that the dataset obtained from test takers' responses to the PISA 2012 reading test questions is close to the simple structure. Item parameters indicate that there is one dominant dimension on which most of the test items have higher discriminations. Three items that bifactor model provided high discrimination parameters both on the general and the sub-dimension had low discrimination parameters based on two parameter logistic model. Item level fit statistics evidenced that the bifactor model provides better item level fit than the two-parameter logistic model. However, there were trivial differences among item fit statistics provided by the models. Item level fit statistics support the results of comparisons made between models based on item parameters. Index values show that the general dimension

explains % 63 of variance explained by the bifactor model, while sub-dimensions representing reading text-effects explain very low and negligible variances. In addition, it was found that the sub-dimensions have very low omega reliability coefficients when the effect of the general dimension on sub-dimensions is controlled.

Discussion and Conclusion

Item discrimination parameters estimated by the bifactor model evidenced that general item discrimination parameters are similar to the item parameters provided by the two-parameter logistic model. The bifactor model provided improvement on the item level fit over the two-parameter logistic model, however, this improvement was trivial. Both models produced similar results in terms of the item level model data fit. Furthermore, the general dimension explained most of the variance observed in the data. Based on these findings, it was concluded that the general dimension representing reading comprehension skill is the dominant dimension underlying the data. Although, reading texts also affect test takers' item responses, a great amount of items better discriminate students in terms of their reading comprehension skills. Therefore, it was concluded that the text effect is small enough to accept that the data holds (essential) unidimensionality assumption. The bifactor approach provided a practical way to determine whether fitting a unidimensional model is harmful enough to bias the item and ability estimations. Therefore, researchers are recommended to analyze dimensionality of the data and the magnitude of reading text effect on item responses when they have a dataset obtained by applying passage-based test items on test-takers.

Giriş

Test geliştirme sürecinde, test maddeleriyle bireylerin etkileşimi sonucunda oluşan madde yanıt matrisinin istatistiksel yapısının incelenmesi önem taşımaktadır. Bu nedenle, psikometri alanında boyutluluk kavramı ve boyutluluk analizleri üzerinde yapılmış çok sayıda çalışma bulunmaktadır. Geleneksel Madde Tepki Kuramı (MTK) çerçevesinde boyutluluk, monotonik ve yerel bağımsız bir model oluşturmak için gereken örtük özellik sayısı olarak tanımlanmaktadır. Boyutluluğun bir diğer tanımı ise yaklaşık tek boyutluluk kavramına odaklanmaktadır. Bu tanımda boyutluluk, zayıf yerel bağımsız bir model oluşturmak için gerekli boyut sayısı olarak ele alınmaktadır (Stout, 1990; Tate, 2003; Zhang, 2016).

Boyutluluk analizleri çoğunlukla ölçme aracının uygulanmasıyla elde edilen veri setinin tek boyutluluk varsayımını karşılayıp karşılamadığını incelemeye odaklanmaktadır. Ancak, bulgular çok boyutluluğun varlığına işaret ettiğinde madde yanıtlarını etkileyen örtük boyutların sayısının ve doğasının ortaya koyulması gerekmektedir. Bu durumda, boyutluluk analizinin temel amaçlarından biri madde yanıt matrisinin karmaşık yapısının belirlenmesi olmaktadır. Örneğin, aynı okuma metnine dayalı madde takımlarını içeren okuduğunu anlama başarısını ölçen bir test verisinin çok boyutlu bir yapıya sahip olması beklenebilir. Böyle bir testte yer alan maddelere bireyler tarafından verilen yanıtlar hem okuduğunu anlama başarısından hem de ait oldukları okuma metninin özelliklerinden etkilenebilmektedir. Bu nedenle testte yer alan her madde, okuduğunu anlama başarısını yansıtan genel boyutla birlikte ait oldukları metinle ilişkili çok boyutlu bir yapıya sahip olabilmektedir (Tate, 2003).

Okuduğunu anlamının değerlendirilmesinde, aynı okuma metnine dayalı farklı maddeler yaygın olarak kullanılmaktadır. Maddeler aynı metne bağlı olduklarında, yanıtlayıcının bir maddeye verdiği yanıtın aynı madde takımı içerisinde yer alan diğer maddelere verdiği yanıtları etkilemesi mümkün olmaktadır (So, 2010). Örneğin, bireyin ilgili metnin konusuna ilişkin bilgi ve deneyimleri sınırlı ise bu o metne dayalı geliştirilmiş tüm maddelerdeki performansını etkileyebilir. Bu durum, tüm test uygulamalarının temelinde yer alan klasik ya da modern test kuramlarının tek boyutluluk ve yerel bağımsızlık gibi en temel varsayımlarını ihlal etmektedir. MTK çerçevesinde tanımlanan tek boyutluluk varsayımı, tek bir örtük boyutun madde yanıtları arasındaki bağımlılığı açıklayabilir olmasını gerektirmektedir (De Ayala, 2009; Hambleton ve Swaminathan, 1983). Madde yanıtlarını etkileyen birden fazla örtük boyut söz konusu olduğunda ise madde yanıt verisi tek boyutluluktan uzaklaşarak karmaşık ve çok boyutlu bir yapıya sahip olmaktadır.

Aynı metne dayalı madde takımlarını içeren bir test kullanıldığında, madde yanıtları birden fazla örtük boyuttan etkilenebileceğinden verinin boyutluluğunun incelenmesi daha da önemli olmaktadır. Böyle bir durumda test maddeleri, yanıtlayıcının maddeye doğru yanıt verebilmek için birden fazla beceriye sahip olmasını gerektirmektedir. Bu testten elde edilen verinin boyutluluğunun tek boyutlu ölçme modellerine dayalı olarak incelenmesi yanıtlayıcıların yeteneklerine ilişkin yanlış çıkarımlarda bulunmaya neden olabilmektedir (Walker ve Beretvas, 2003; Wang, Chen ve Cheng, 2004). Deng, Wells ve Hambleton'a (2008) göre, yaklaşık tek boyutluluğun varlığını doğrulamadan parametre kestiriminde MTK modellerini uygulamak, kuramın üstün yönlerinden faydalanmayı engellemekte ve hatalı sonuçlara neden olmaktadır. Ackerman (1994) da çok boyutlu yanıt matrisinde tek boyutlu modellerin uygulanmasının MTK'nın değişmezlik özelliğini olumsuz etkileyebileceğini ve yanıtların doğasına ilişkin yanlış sonuçlar çıkarmaya yol açabileceğini ifade etmektedir. Bu nedenle, kullanılacak ölçme modeline karar vermeden önce veri matrisinin boyutluluğunun incelenmesi gerekmektedir.

Boyutluluğun incelenmesinde kullanılan farklı yöntemler verinin yapısına ilişkin farklı bulgular sağlayabilmektedir. Aynı metne bağlı madde takımlarından oluşan bir okuduğunu anlama başarı testinden elde edilen verinin boyutluluğuna ilişkin farklı yöntemler sonucunda çelişkili bulgular elde edildiğinde, iki faktör modeli (İFM) tek boyutluluk varsayımında bulunma ya da reddetme kararını vermede kullanılabilir önemli bilgiler sağlamaktadır. İFM'de her madde için biri genel boyut, diğeri de ait olduğu alt boyutta olmak üzere iki ayırt edicilik parametresi kestirilmektedir. Bu şekilde İFM, genel boyutun maddeler arasındaki ilişkileri açıklama gücünü ortaya koyarak tek boyutluluk varsayımının incelenmesini sağlamaktadır. Ayrıca, tek boyutlu model ile kestirilen madde ayırt edicilik parametrelerinin İFM ile maddeler için genel boyutta kestirilen parametrelerle karşılaştırılması yoluyla tek boyutluluktan uzaklaşma derecesine ilişkin de bilgi elde edilmektedir. Tek boyutlu model ve İFM ile kestirilen ayırt edicilik parametrelerinin birbirine benzer olması madde yanıt verisinin tek boyutlu yapıdan çok fazla uzaklaşmadığını göstermektedir (Reise, Morizot, & Hays, 2007; Luo, & Al-Harbi, 2016). İFM'nin sağladığı bu önemli bilgilere rağmen, özellikle Türkçe alan yazındaki boyutluluk çalışmalarında kullanımı çok sınırlıdır. Bu nedenle, boyutluluğun incelenmesinde İFM'nin kullanılmasına yol gösterici olacak çalışmaların yapılmasına gereksinim duyulmaktadır. Özellikle aynı metne dayalı maddelerden elde edilen yanıt verisinin boyutluluğunun dikkatle incelenmesinin önemli olduğu düşünülmektedir. Çünkü bu tür maddelerin MTK'nın tek boyutluluk ve yerel bağımsızlık gibi temel varsayımlarını ihlal etme olasılığı yüksektir. Bu nedenle, araştırmanın amacı PISA 2012 Türkiye örnekleminde yer alan öğrencilerin okuduğunu anlama test

maddelerine verdikleri yanıtlardan elde edilen verinin boyutluluğunun iki parametrelili lojistik model (2PLM) ve İFM'ye dayalı olarak incelenmesidir.

Yöntem

Araştırmanın Deseni

Bu araştırma, 2PLM ve İFM'ye dayalı olarak yapılan madde parametre kestirimlerinin ve madde düzeyinde model veri uyumlarının karşılaştırılması yoluyla madde yanıt verisinin boyutluluğunun incelenmesini amaçlayan nicel bir araştırmadır.

Çalışma Grubu

Araştırmanın çalışma grubunu PISA 2012 uygulamasında 12 numaralı test formunun uygulandığı öğrenciler oluşturmaktadır. Bu forma yanıt veren öğrenciler içerisinde kayıp veri içermesi nedeniyle iki öğrenci çıkarıldıktan sonra araştırma grubunda 284 öğrenci yer almıştır. Graham (2009, s. 554) tarafından belirtildiği gibi kayıp veri içermesi nedeniyle kaybedilen yanıtlayıcı oranının % 5'i geçmemesi durumunda, veri silme yöntemi yanlı parametre kestirimlerine neden olmamaktadır. Bu araştırmanın çalışma grubunda kayıp veri nedeniyle kaybedilen öğrenci oranı yalnızca % 0.69 (iki öğrenci) olduğundan kayıp veri silme yöntemi kullanılmıştır. Araştırmanın çalışma grubunu oluşturan öğrencilerin tamamı 15 yaşındadır. Öğrencilerin cinsiyete göre dağılımları incelendiğinde 284 öğrenciden 145'inin kız, 139'unun erkek öğrenci olduğu görülmektedir. Öğrencilerin okuduğunu anlama testi maddelerine verdikleri yanıtlardan elde edilen veri üzerinde normallik ve uç değer açısından incelemeler yapılmıştır. Veri, uç değerler açısından incelenirken z puanı ± 3.29 aralığını ölçüt olarak kabul edilmiştir (Tabachnick ve Fidell, 2001). Öğrencilerin z puanları -2.38 ve 2.29 arasında değiştiğinden, hiç bir öğrencinin veri setinden çıkarılması gerekmemiştir. Madde puanlarının ve testin tamamından elde edilen toplam puanların normallik varsayımı açısından uygunluğu çarpıklık ve basıklık değerlerine dayalı olarak incelenmiştir. Toplam puanlar için çarpıklık katsayısı -0.18, basıklık katsayısı -0.51 olarak hesaplanmıştır. Kline'a (2011, s. 63) göre normallik varsayımının sağlandığının kabul edilmesi için çarpıklık ve basıklık değerlerinin 3'ten küçük olması gerekmektedir. Buna göre, verinin normal dağılıma sahip olduğu görülmüştür.

Veri Toplama Araçları

PISA 2012 uygulamasında 13 farklı test formu kullanılmıştır. 13 farklı form içerisinde en çok sayıda öğrenciye uygulanmış olması nedeniyle 12 numaralı formdan elde edilen verinin kullanılmasına karar verilmiştir. Bu formda *Uyku ile Çikolata ve Sağlık* başlıklı metinlere ait dörder madde, *Kokeshi Bebekleri ile Dürüst Ticaret* başlıklı metinlere ait üçer madde olmak üzere toplamda 14 madde yer almaktadır. Bu araştırma kapsamında *Uyku* metni *Metin 1*, *Kokeshi Bebekleri* *Metin 2*, *Dürüst Ticaret* *Metin 3*, *Çikolata ve Sağlık* *Metin 4* biçiminde kodlanmıştır.

Veri Analizi

PISA 2012 okuduğunu anlama testi 12 numaralı formda yer alan 14 maddeye ilişkin parametre kestirimleri ve madde uyum istatistikleri MTK çerçevesinde İFM ve 2PLM'ye dayalı olarak yapılmıştır. Tüm parametre ve model veri uyumu kestirimleri açık kaynaklı bir istatistik programı olan R programında "mirt" (Chalmers, 2012) paketi kullanılarak yapılmıştır.

2PLM'de, maddeler için biri güçlük (b-parametresi) diğeri ayırt edicilik (a-parametresi) olmak üzere iki parametre kestirilmektedir. Güçlük parametresi, madde güçlüğü'nün bir ölçüsünü vermektedir. Madde güçlük parametresi için elde edilen yüksek değerler maddenin zor bir madde olduğunu; düşük değerler ise kolay bir madde olduğunu göstermektedir. Güçlük parametresi, örtük özellik ölçeği üzerinde madde karakteristik eğrisinin yerini belirlemektedir. Madde zorlaştıkça eğri soldan sağa doğru hareket etmektedir. Ayırt edicilik parametresi, madde yanıt fonksiyonunun maksimum eğim değerini vermektedir. Madde yanıt fonksiyonunun eğimi, doğru yanıt olasılığının 0.50 olduğu θ düzeyinde en dik olmaktadır. Dolayısıyla, madde güçlük parametresi fonksiyonunun eğiminin en dik olduğu noktayı, ayırt edicilik parametresi ise bu noktadaki eğimin değerini göstermektedir. Ayırt edicilik parametresinin değeri arttıkça, maddedeki doğru yanıt olasılığı ile örtük özellik arasındaki ilişki de artmaktadır. Bu nedenle, a-parametresi yüksek olan bir madde farklı örtük özellik düzeyine sahip olan bireyleri güçlü biçimde ayırt edebilmektedir (Reckase ve McKinley, 1991; Ostini ve Nering, 2006; Liu, 2007; Reckase, 2009).

İFM, çok düzeyli yapıyı modellemeyi sağlayan parametrik bir çok boyutlu MTK modelidir. İFM'ye dayalı kestirimler yapısal eşitlik modellemesine ve MTK'ya dayalı olarak iki farklı çerçevede yapılabilmektedir (Thissen ve Wainer, 2001; Brown, 2006; Berkeljon, 2012; Qinn, 2014). Çok boyutlu MTK çerçevesinde ikili puanlanan maddelerde kullanılacak iki parametrelili İFM'nin lojistik fonksiyonu Formül 1'de verilmiştir.

$$P(X_i = 1 | \theta_j, a_i, d_i) = \frac{1}{1 + e^{-(a_{iG}\theta_G + a_{iS}\theta_S + d_i)}} \quad \text{Formül 1}$$

Formül 1, İFM'ye dayalı olarak ikili puanlanan bir maddedeki doğru yanıt olasılığının fonksiyonunu göstermektedir. Formülde, θ_G bireyin genel boyutta, θ_S ise k sayıda alt boyutlardan birinde sahip olduğu örtük özellik düzeyini göstermektedir. Ayrıca, a_{iG} ve a_{iS} sırasıyla maddenin genel boyut ve alt boyutlardan birindeki ayırt edicilik parametrelerini ifade etmektedir. Formülde verilen model fonksiyonundan da görülebileceği gibi herhangi bir madde için biri genel boyutta diğeri alt boyutlardan birinde olmak üzere iki ayırt edicilik parametresi kestirilmektedir. Çok boyutlu MTK modellerinde olduğu gibi d_i maddenin güçlüğü ile ilgili parametreyi ifade etmektedir (DeMars, 2006; Reckase, 2009).

İFM ile kestirilen madde parametrelerine dayalı olarak açıklanan ortak varyans (ECV) değeri ve omega güvenilirlik katsayıları hesaplanmıştır. Bunun için İFM ile kestirilen ayırt edicilik parametreleri standartlaştırılmış faktör yüklerine dönüştürülmüştür. Genel boyut ve alt boyutlar için ECV değeri, maddeler için genel boyutta ya da ilgili alt boyutta kestirilen faktör yüklerinin karelerinin toplamının, genel ve alt boyutlarda kestirilen faktör yüklerinin karelerinin toplamına oranı alınarak hesaplanmıştır (Qinn, 2014). Genel ve grup faktörlerin güvenilirliklerinin incelenmesi Omega katsayılarına dayalı olarak yapılmıştır. Alt boyutlar için Omega katsayısı hesaplanırken payda ilgili alt boyutta yer alan maddelerin genel boyut ve alt boyuttaki yüklerinin kareleri toplamı alınmıştır. Paydada ise bu toplama hata da dahil edilmiştir. Maddeler için hata değerleri genel boyut ve alt boyuttaki yüklerinin kareleri toplamının 1'den çıkarılmasıyla elde edilmiştir (Reise, Bonifay ve Haviland, 2013).

Madde düzeyinde model veri uyumu, Orlando ve Thissen (2000) tarafından geliştirilmiş $S-\chi^2$ istatistiği kullanılarak incelenmiştir. Ki-kare testi uygulamak için kestirilen teta değerlerine göre bireyler sıralanarak belirli sayıda alt gruba ayrılmaktadırlar. Her grupta, ilgili yanıtı veren birey sayısı hesaplanmaktadır. Bu sayı daha sonra madde yanıt fonksiyonuna dayalı olarak kestirilen değer ile karşılaştırılmaktadır (Chon, Lee ve Anslye, 2007; Kang ve Chen, 2011, Reise, 1990).

Bulgular

Maddeler için iki modele dayalı olarak kestirilen ayırt edicilik parametreleri Tablo 1'de verilmiştir.

Tablo 1
Madde Ayırt Edicilik Parametreleri

	a_{2PLM}	a_G	a_S	d
Metin 1				
1	1.77	1.71	0.31	1.80
2	1.23	1.68	1.80	0.43
3	1.31	1.24	0.25	-1.70
4	1.16	1.13	0.14	0.00
Metin 2				4.39
1	0.79	2.29	5.49	-1.06
2	0.51	0.49	0.33	2.62
3	1.82	1.75	0.56	-0.45
Metin 3				0.64
1	0.92	1.11	0.61	4.91
2	0.84	0.85	0.29	-0.80
3	1.27	2.74	3.36	2.28
Metin 4				0.38
1	0.87	0.88	0.11	-3.11
2	1.13	1.27	0.57	1.80
3	1.03	2.37	3.07	0.43
4	1.50	1.48	0.41	-1.70

2PLM= 2 parametrelili lojistik model, İFM= İki faktör modeli

a_{2PLM} = 2 parametrelili lojistik model ile kestirilen ayırt edicilik parametresi,

a_G = Genel boyut ayırt edicilik parametresi, a_S = Alt boyut ayırt edicilik parametresi

Tablo 1’de verilen ayırt edicilik parametreleri incelendiğinde, İFM ile test maddeleri için biri genel boyutta karşılık gelen okuduğunu anlama başarısında, diğeri de ait olduğu okuma metnine karşılık gelen alt boyutta olmak üzere iki ayırt edicilik parametresi kestirildiği görülmektedir. 2PLM’de ise maddeler için bir ayırt edicilik parametresi kestirilmiştir. 2PLM kestirimlerine göre maddelerin ayırt edicilik parametreleri 0.58 ile 1.821 arasında değişmektedir. İFM ile maddeler için genel boyutta kestirilen ayırt edicilik parametreleri 0.485 ile 2.737 arasında dağılmıştır. İki model ile maddeler için kestirilen ayırt edicilik parametrelerinin birbirine benzer olduğu görülmektedir. PISA okuduğunu anlama başarı testinin 12 numaralı formunda yer alan 14 madde içerisinde 11 madde için iki modele dayalı olarak kestirilen ayırt edicilik parametreleri arasında 0.45’ten daha büyük bir fark olmadığı bulunmuştur. İFM ile maddeler için kestirilen genel boyut ayırt edicilik parametreleriyle 2PLM ile kestirilen ayırt edicilik parametreleri arasındaki benzerlik, yaklaşık tek boyutluluğun sağlandığına ve verinin yaklaşık basit yapıda olduğuna ilişkin bir kanıt oluşturmaktadır (Reise, Morizot ve Hays, 2007).

İki modele dayalı olarak kestirilen ayırt edicilik parametreleri arasındaki farkın büyük olduğu üç madde için İFM ile kestirilen genel ve alt boyut ayırt edicilikleri incelenmiştir. Buna göre, Kokeshi Bebekleri metnine bağlı maddelerden biri için genel boyutta 2.29, alt boyutta ise 5.49 ayırt edicilik parametreleri kestirilmiştir. Dürüst Ticaret metnine dayalı bir madde için genel ve alt boyut ayırt edicilik parametreleri sırasıyla 2.74 ve 3.36 olarak elde edilmiştir. Çikolata ve Sağlık metnine ait bir maddenin de genel ve alt boyut ayırt edicilik parametreleri sırasıyla 2.37 ve 3.07 olarak kestirilmiştir. Bu üç maddenin İFM’ye dayalı olarak kestirilen genel ve alt boyut ayırt ediciliklerinin yüksek olduğu bulunmuştur. Ancak, tek boyutlu MTK modeli olan 2PLM ile bu üç madde için diğer maddelere göre daha düşük ayırt edicilik parametreleri kestirilmiştir. Tek boyutluluğun karşılanıp karşılanmadığını incelemek amacıyla kestirilen madde uyum istatistikleri Tablo 2’de verilmiştir.

Tablo 2
Madde Uyum İstatistikleri

	2PLM		İFM	
	$S-\chi^2$	p	$S-\chi^2$	P
Metin 1				
1	4.11	0.77	3.83	0.70
2	3.15	0.87	3.09	0.80
3	11.47	0.12	11.17	0.08
4	14.00	0.06	13.77	0.03
Metin 2				
1	10.23	0.25	9.99	0.19
2	6.90	0.65	7.14	0.52
3	7.11	0.21	7.41	0.12
Metin 3				
1	2.86	0.94	2.74	0.91
2	7.20	0.51	7.03	0.43
3	5.01	0.66	5.09	0.53
Metin 4				
1	9.70	0.29	9.83	0.20
2	6.67	0.46	6.56	0.36
3	15.18	0.03	14.16	0.03
4	4.13	0.53	4.05	0.40

Tablo 2’de verilen uyum istatistikleri incelendiğinde 2PLM’ye dayalı olarak maddeler için kestirilen değerlerin 2.86 ile 15.18 arasında değiştiği görülmektedir. İFM’ye göre kestirilen madde uyum istatistikleri ise 2.75 ile 14.16 arasında dağılmıştır. Madde uyum istatistiklerine dayalı olarak hesaplanan anlamlılık değerlerine göre testte yer alan maddelerden yalnızca biri iki modele de uyum sağlamamıştır. Bunun dışında testte yer alan 14 maddeden 13’ü tek boyutlu modele, 12’si ise İFM’ye uyum sağlamıştır. İFM’ye dayalı olarak tek boyutlu modele göre genel olarak daha düşük istatistikler kestirilmiştir. İFM ile hesaplanan madde uyum istatistiklerine göre, 12 madde içerisinde 8 madde için 2PLM ile kestirilenden daha iyi uyum istatistikleri elde edilmiştir. Ancak, iki model ile kestirilen değerler arasındaki farklar göz ardı edilebilecek kadar küçüktür. Modellerin maddeler için birbirine çok yakın uyum istatistikleri sağladığı ifade edilebilir. Madde uyum istatistikleri de öğrencilerin madde yanıtlarının modellenmesinde tek boyutlu bir modelin kullanılabilmesini ve yaklaşık tek boyutluluğun sağlandığını desteklemektedir. Tek boyutun madde yanıtlarını açıklamadaki gücünü incelemek

amacıyla açıklanan ortak varyans (ECV) değeri, alt ölçek omega ve hiyerarşik omega katsayıları hesaplanmıştır. Tablo 3'te genel boyut ve alt boyutlar için elde edilen değerler verilmiştir.

Tablo 3
Açıklanan Ortak Varyans Değerleri ve Omega Katsayıları

	Genel Boyut	Metin 1	Metin 2	Metin 3	Metin 4
ECV	0.63	0.06	0.13	0.09	0.09
ω		0.94	0.91	0.91	0.93
ω_H	0.79	0.02	0.07	0.05	0.04

ECV = Açıklanan ortak varyans değeri,
 ω = Omega katsayısı,
 ω_H = Hiyerarşik omega katsayısı

Tablo 3'te genel boyut ve alt boyutlar için hesaplanan ECV değerleri incelendiğinde, bu değerlerin 0.06 ile 0.63 arasında değiştiği görülmektedir. Dört okuma metninin madde yanıtlarını açıklamadaki gücünü gösteren ECV değerleri ise 0.06 ile 0.13 arasında değişmiştir. Genel boyut için ECV değeri 0.63 olarak hesaplanmıştır. Uyku metni için bu değer 0.06, Kokeshi Bebekleri için 0.13, Dürüst Ticaret ile Çikolata ve Sağlık metinleri için de 0.09'dur. İFM ile açıklanan varyansın % 63'ü genel boyut tarafından açıklanmaktadır. Alt boyutlar tarafından oldukça düşük varyans açıklanmaktadır. Reise, Scheines, Widaman ve Haviland'a (2013) göre genel boyut için 0.60'ın üzerinde bir ECV değeri elde edilmesi, veri yapısının yaklaşık tek boyutluluğu sağladığına ilişkin bir kanıt oluşturmaktadır. Genel boyut için elde edilen değer 0.60'ın üzerinde olması baskın bir tek boyutun varlığına işaret etmektedir. Buna göre, genel boyut için yüksek ECV değeri elde edilmiş olması bireylerin madde performanslarının çok büyük oranda okuduğunu anlama başarısından etkilendiğini göstermektedir. Alt boyutlar için kestirilen değerlerin çok düşük olması, veride maddelerin bağlı oldukları metnin özelliklerinden kaynaklı varyansın çok düşük olduğunu göstermektedir.

Omega katsayısı alt boyutların güvenilirliklerine ilişkin bilgi vermektedir. Alt boyutlar için kestirilen katsayılar 0.91 ve 0.94 arasında değişmektedir. Omega katsayılarının yüksek olması, alt boyutların güvenilirliklerinin yüksek olduğu anlamına gelmektedir. Hiyerarşik omega katsayısı veride ilgili boyuta atfedilebilir toplam varyansa ilişkin bilgi vermektedir (Periard, 2016). Bu katsayı genel boyut için 0.79 olarak hesaplanmıştır. Bu testten elde edilen toplam puan varyansının % 79'unun okuduğunu anlama başarısı açısından bireyler arasındaki farklılıklardan kaynaklandığını göstermektedir. Alt boyutların omega katsayıları oldukça yüksek iken genel boyutun etkisi çıkarılarak hesaplanan hiyerarşik omega katsayıları çok düşüktür. Alt ölçekler için hesaplanan hiyerarşik omega katsayıları 0.02 ile 0.07 arasında değişmektedir. Bu katsayılar, okuduğunu anlama başarısının etkisi çıkarıldığında metne ilişkin özellikler ile açıklanan güvenilir varyansın çok düşük olduğunu göstermektedir.

Sonuç ve Tartışma

İFM'ye dayalı olarak kestirilen madde parametreleri, testte yer alan maddelerin büyük çoğunluğunun alt boyutlara göre genel boyutta daha yüksek ayırt edicilik parametrelerine sahip olduklarını göstermiştir. Ayrıca, genel boyutta kestirilen ayırt edicilik parametrelerinin 2PLM'ye dayalı olarak kestirilen ayırt edicilik parametreleriyle benzer olduğu bulunmuştur. Madde parametreleri arasında elde edilen bu benzerlik, öğrencilerin aynı metinlere dayalı olan madde takımlarına verdikleri yanıtlardan elde edilen verinin yaklaşık basit yapıda olduğunu göstermektedir. Öğrencilerin maddelere verdikleri yanıtların altında yatan baskın bir boyut bulunmaktadır. Ayrıca, modeller madde düzeyinde model veri uyumu açısından benzer sonuçlar sağlamıştır. Çok boyutlu model, tek boyutlu modele göre madde düzeyinde model veri uyumunda önemli bir iyileşme yaratmamıştır. Bu da bireylerin madde performansları altında yatan tek ve baskın bir boyutun varlığını desteklemektedir.

İFM ile kestirilen madde parametrelerine dayalı olarak genel boyut için yüksek bir ECV değeri elde edilirken, alt boyutlar için hesaplanan değerlerin düşük olduğu görülmüştür. Ayrıca, genel boyutun etkisi çıkarıldığında alt boyutların güvenilirliklerinin çok düşük olduğu bulunmuştur. Alt boyutlar için hesaplanan değerlerin çok düşük olması, okuma metinlerinin öğrencilerin yanıtlarında okuma metninin özelliklerinden kaynaklanan farklılaşmanın çok düşük ve önemsiz olduğunu göstermektedir. Okuma metninin öğrencilerin madde yanıtları üzerinde bir miktar etkisi olsa da maddeler okuduğunu anlama becerisi açısından öğrenciler arasındaki farklılıkları daha iyi ayırt etmektedir. Bu nedenle, okuma metninin öğrenci yanıtları üzerindeki etkisinin, test verisinin yaklaşık tek boyutluluğu sağladığını kabul etmeye izin verecek kadar düşük olduğu sonucuna varılmıştır. Genel ve alt boyut ayırt edicilikleri çok yüksek olan üç madde için tek boyutlu modelle daha düşük ayırt edicilik parametreleri kestirilmiştir. Bu bulguya dayalı olarak çok boyutluluğun tek boyutlu modele dayalı olarak yapılan madde parametre kestirimlerini etkileyebileceği düşünülmüştür. Bu bulguya benzer şekilde,

araştırmacılar da çok boyutlu veri matrisi üzerinde tek boyutlu ölçme modelleri kullanıldığında, madde ve birey parametre kestirimlerinin ikincil boyuttan etkilendiğini ifade etmektedir (DeMars, 2006; Drasgow ve Parsons, 1983).

Bu araştırmanın belirli sınırlılıkları bulunmaktadır. Araştırma, PISA 2012 Türkiye uygulamasında, okuduğunu anlama başarısını ölçen test formlarından yalnızca 12 numaralı formun uygulandığı yanıtlayıcı grubundan elde edilen veri seti üzerinde yürütülmüştür. Çok boyutlu MTK uygulamalarının doğru parametre kestirimleri elde etmek için geniş örneklem üzerinde yürütülmesi gerektiği düşünüldüğünde, bu durum örneklem büyüklüğü açısından önemli bir sınırlılık getirmiştir. PISA 2012 uygulamasına katılan ülkeler okuma becerisi açısından farklı özelliklere sahip olabileceklerinden, bu ülkelerden elde edilen veri setlerinin birleştirilerek örneklem büyüklüğünün artırılması uygun görülmemiştir. Bu nedenle, bu araştırmanın iki faktör modelinin verinin boyutluluğunun incelenmesinde nasıl kullanılacağına ilişkin yol gösterici bir ön çalışma olarak görülmesi gerekmektedir. Araştırmanın bir diğer sınırlılığı PISA uygulamasında kullanılan test maddelerinin gizlilik ilkesi nedeniyle yayınlanmamasıyla ilgilidir. Bu durum, madde düzeyinde yapılan incelemelerin sınırlı kalmasına, boyutların temsil ettiği özelliklere ilişkin yalnızca istatistiksel bulgulara dayalı olarak yorum yapılmasına neden olmuştur. Bu sınırlıklar göz önünde bulundurularak, iki faktör modelinin veri seti üzerinde tek boyutlu model kullanılmasının madde kestirimlerini etkileme düzeyi ile aynı metne dayalı madde takımlarından elde edilen verinin boyutluluğunu incelemeye pratik bir yol sağladığı ifade edilebilir. Bu nedenle, benzer çalışmalar yürütecek araştırmacılara ve uygulayıcılara, aynı okuma metnine dayalı madde takımları içeren daha geniş örneklemelerden elde edilen, daha fazla sayıda okuma metninin kullanıldığı ve madde düzeyinde incelemelere izin verecek veri setleri üzerinde boyutluluğun ve okuma metninin yanıtlar üzerindeki etkisinin incelenmesinde İFM kullanmaları önerilmektedir.

Kaynakça/References

- Ackerman, T.A. (1994). Using multidimensional item response theory to understand what items and tests are measuring, *Applied Measurement in Education*, 7, 255-278.
- Berkeljon, A. (2012). *Multidimensional item response theory in clinical measurement: A bifactor graded-response model analysis of the outcome questionnaire-45.2*. (Unpublished Dissertation). Brigham Young University, Utah.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- Chalmers, R. P. (2012). A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1-29.
- Chon, K. H., Lee, W., & Ansley, T. N. (2007). *Assessing IRT model-data fit for mixed format tests*. Iowa: Center for Advanced Studies in Measurement and Assessment.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.
- DeMars, C. (2006). *Item response theory*. New York: Oxford University Press, Inc.
- Deng, N., Wells, C., & Hambleton, R. (2008). *A confirmatory factor analytic study examining the dimensionality of educational achievement tests*. Connecticut: NERA Conference Proceedings.
- Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement*, 7, 189-199.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annu. Rev. Psychol.*, 60, 549-576.
- Hambleton, R. K., & Swaminathan, H. (1983). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Kang, T., & Chen, T. T. (2011). Performance of the generalized S- χ^2 item fit index for the graded response model. *Asia Pacific Educ. Rev.*, 12, 89-96.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. NY: The Guilford Press.
- Liu, J. (2007). *Comparing multidimensional and unidimensional computer adaptive strategies in psychological and health assessment*. (Unpublished Dissertation). University of North Carolina at Chapel Hill, North Carolina.
- Luo, Y., & Al-Harbi, K. (2016). The Utility of the Bifactor Method for unidimensionality assessment when other methods disagree: an empirical illustration. *SAGE Open*, 1-7.
- Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24, 50-64.
- Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models*. California: Sage Publications.
- Periard, D. (2016). A bifactor model of burnout? *An item response theory analysis of the Maslach burnout inventory*. (Unpublished Dissertation). Wright State University, Dayton.
- Quinn, H. O. (2014). *Bifactor models, explained common variance (ECV) and the usefulness of scores from unidimensional item response theory analyses*. (Unpublished Master's Thesis). University of North Carolina at Chapel Hill, North Carolina.
- Reckase M. D., & McKinley, R. L. (1991) The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement*, 15, 361-373.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer.
- Reise, S. P. (1990). A comparison of item- and person-fit methods of assessing model-data fit in IRT. *Applied Psychological Measurement*, 14(2), 127-137.
- Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, 95(2), 129-140.
- Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual Life Res*, 16, 19-31.

- Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement, 73*(1), 5-26.
- So, Y. (2010). *Dimensionality of responses to a reading comprehension assessment and its implications to scoring test takers on their reading proficiency*. (Unpublished Dissertation). University of California, Los Angeles.
- Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika, 55*, 293-325.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. MA: Allyn & Bacon, Inc.
- Tate, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological Measurement, 27*, 159–203.
- Thissen, D., & Wainer, H. (2001). *Test scoring*. NJ: Lawrence Erlbaum Associates, Inc.
- Walker, C. M., & Beretvas, S. N. (2003). Comparing multidimensional and unidimensional proficiency classifications: Multidimensional IRT as a diagnostic aid. *Journal of Educational Measurement, 40*(3), 255-275.
- Wang, W., Chen, P., & Cheng, Y. (2004). Improving measurement precision of test batteries using multidimensional item response models. *Psychological Methods, 9*(1), 116-136.
- Zhang, M. (2016). *Exploring dimensionality of scores for mixed-format tests*. (Unpublished Dissertation). University of Iowa, Iowa.