

PISA 2009 TUTUM ANKETİ MADDE PUANLARININ AŞAMALI MADDE TEPKİ MODELİ İLE İNCELENMESİ

Arş. Gör. Şeyma Uyar
Mehmet Akif Ersoy Üniversitesi, Burdur
syuksel@mehmetakif.edu.tr

Arş. Gör. Neşe Öztürk Gübeş
Hacettepe Üniversitesi, Ankara
neseozturk@hacettepe.edu.tr

Prof. Dr. Hülya Kelecioğlu
Hacettepe Üniversitesi, Ankara
hulyaebb@hacettepe.edu.tr

Özet

Bu araştırmada, Uluslararası Öğrenci Değerlendirme Programı (PISA) 2009 kapsamında uygulanan okumaya yönelik tutum anketinden elde edilen madde puanlarının aşamalı tepki modeli (Graded Response Model (GRM)) ile incelenmesi amaçlanmaktadır. Araştırma, Türkiye örneğinde maddelerin tamamına yanıt veren 4198 öğrencinin cevapları kullanılarak yürütülmüştür. Tutum anketinde 11 madde bulunmaktadır ve bu maddelere öğrencilerin “hiç katılmıyorum (1)” ve “tamamen katılıyorum (4)” arasında tepkide bulunmaları istenmiştir. Tutum maddeleri, çok kategorili madde tepki kuramı modellerinden likert tipi ölçekler için uygun olan aşamalı tepki modeli kullanılarak incelenecektir. Veri analizinin ilk aşamasında verilerin madde tepki kuramı varsayımlarını sağlayıp sağlamadığı incelenmiş; ikinci aşamada ise MULTILOG 7.03 bilgisayar programında aşamalı tepki modeli ile madde kalibrasyonu yapılmıştır. Araştırmanın sonucunda, model-veri uyumunun sağlandığı görülmüştür. Elde edilen bulgulara göre; okumaya yönelik tutumları açısından öğrencileri en iyi ayırt eden dolayısıyla en çok bilgi veren maddenin 3. madde, en kötü ayırt eden ve en az bilgiyi veren maddenin ise 9. madde olduğu sonucuna ulaşılmıştır.

Anahtar Sözcükler: Madde Tepki Kuramı, Aşamalı Tepki Modeli, Tutum

EXAMINING THE PISA 2009 ATTITUDE QUESTIONNAIRE ITEM SCORES WITH USING GRADED RESPONSE MODEL

Abstract

In this study, the item scores obtained from “Attitude to Reading” part of PISA 2009 Student Survey was examined in respect of the graded response model. The study was carried out by means of 4198 data in Turkey sample, so the answers of 4198 respondents are considered. There are 11 items in the survey and 4 response categories an ordinal score ranging from 1 (strongly not agree) to 4 (strongly agree) is assigned to each examinee for each item. The items of attitude, was calibrated with the Graded Response Model, which is appropriate to Likert type data. In the first step, the data was examined whether the assumptions of IRT model was provided. Next step, we calibrated the items using GRM in MULTILOG 7.03. Then, item discrimination and difficulty parameters were examined for identifying items quality. According to results, seen that model has good fit to data. Findings showed that; item 3 has the highest slope parameter and gives the maximum information, item 9 has the lowest slope parameter and gives the minimum information about students’ attitude to reading.

Key Words: Item Response Theory, Graded Response Model, Attitude.

GİRİŞ

Türkiye, ulusal boyutta yapılan öğrenci başarısını belirleme çalışmalarını uluslararası boyutta da sürdürmek, kendi öğrencilerinin başarı düzeylerini diğer ülkeler ile karşılaştırarak eğitim sisteminin güçlü ve iyileştirmeye açık yönlerini belirlemek için uluslararası çalışmalara katılmaktadır. Bu çalışmalardan biri de Uluslararası Öğrenci Değerlendirme Programı (PISA)'dır. PISA, Ekonomik İşbirliği ve Kalkınma Teşkilatı-OECD (Organisation for Economic Cooperation and Development) tarafından üç yılda bir düzenlenen; 15 yaş grubu öğrencilerin matematik, fen ve okuma becerileri alanlarındaki bilgi ve becerilerinin değerlendirildiği en büyük eğitim araştırmalarından biridir. PISA projesinde, öğrencilerin üç temel konu alanındaki bilgi ve becerilerini değerlendirmenin yanında öğrencilerin öğrenme stratejileri, problem çözme becerileri ve değerlendirilen konu alanına yönelik ilgi ve tutumları da araştırılmaktadır. PISA uygulamasında temel olarak 13 adet kitapçık, 1 adet öğrenci anketi ve 1 adet okul anketi kullanılmaktadır. PISA 2009'da ağırlıklı olarak temel alınan konu okuma becerileri olduğu için öğrencilerin okumaya yönelik tutumları ve öğrenme stratejileri de değerlendirilmiştir (Milli Eğitim Bakanlığı [MEB], 2010)

PISA uygulamalarında kullanılan çoğu ölçek Madde Tepki Kuramı (MTK) Ölçekleme yöntemleri kullanılarak değerlendirilmektedir. Eğitimde ve psikolojide test geliştirme çalışmalarında madde ve test istatistiklerini kestirmek için MTK'nın yanı sıra klasik test kuramından da yararlanılmaktadır. Hatta KTK uygulamalarına daha sık rastlanmaktadır. KTK birçok test durumuna uygulanabilmesi ve zayıf teorik temelleri olmasından dolayı daha çok tercih edilmektedir (Hambleton ve Jones, 1993). KTK'nın odaklandığı temel durum madde güçlüğü ve ayırt edicilik parametresidir (Doğan ve Tezbaşaran, 2003). Örtük özellikleri ölçmek için kullanılan ölçme araçlarında da temelde maddelerin birbiriyle ve ölçmek istediği özellik ile ilişkisi incelenir. MTK, KTK' dan farklı olarak bireyin yeteneğini de işin içine katarak maddelerden elde edilen bilginin belli yetenek düzeylerine göre ele alınması konusunda avantaj sağlar. Hambleton ve Swaminathan (1985), bir madde tepki modelini "bireyin gözlenen test performansı ile bu test performansının altında yatan gözlenemeyen örtük özellik ya da yetenekler arasındaki ilişkiyi tanımlayan model" (s.9) olarak ifade etmişlerdir. Madde tepki kuramında bireyin yetenek düzeyi maddeye verdiği yanıt aracılığı ile kestirilir. Bu nedenle madde tepki kuramı modelleri, bireyin maddeye verdiği yanıtlar ile bireyin yetenek düzeyi ve maddenin özelliklerini ilişkilendirir. Söz konusu duruma göre madde tepki kuramı, model temelli ölçmelere dayanır (Emretson ve Reise, 2000).

MTK'de ikili ve çoklu puanlanan maddeler için çeşitli modeller geliştirilmiştir. Her ne kadar yapılan çalışmaların büyük bir çoğunluğu ikili puanlanan maddelerden elde edilen veri analizine yoğunlaşmış olsa da ikiden fazla puan kategorisine sahip maddeler için de önemli modeller bulunmaktadır. Aşamalı tepki modeli (Graded Response Model (GRM); Samejima, 1969), kısmi puan modeli (Partial Credit Model (PCM); Masters, 1982) ve genelleştirilmiş kısmi puan modeli (Generalized Partial Credit Model (GPCM); Muraki, 1992) en çok çalışılan ve çok boyutlu MTK çerçevesinde genellemesi yapılan üç modeldir (Reckase, 2009). Bu araştırmanın temel amacı PISA 2009 kapsamında uygulanan okumaya yönelik tutum ölçeği maddelerini aşamalı madde tepki modelini kullanarak incelemektir.

Aşamalı tepki modeli (ATM), Likert ölçeği gibi sıralı kategorilere sahip çoklu puanlanan maddeler için uygun bir modeldir ve 2 PL (iki parametrelili lojistik model) modelin genelleştirilmiş bir versiyonu olarak düşünülebilir. Model, bütün maddelerin eşit sayıda kategoriye sahip olmasını gerektirmez. Modelde, her bir ölçek maddesi (i); bir madde eğim parametresi (a_i) ve m_i tane kategoriler arası eşik parametreleri (β_{ij}) ile tanımlanır. m_i , bir maddenin kategori sayısının bir eksiğine eşittir, $m_i = k_i - 1$ 'dir. Aşamalı tepki modeli, matematiksel olarak (1) numaralı eşitlikte olduğu gibi ifade edilebilir (Emretson ve Reise, 2000):

$$P_{ik}^*(\theta) = \frac{\exp(a_i(\theta - \beta_{ij}))}{1 + \exp(a_i(\theta - \beta_{ij}))} \quad (1)$$

$$x = j = 1, \dots, m_i$$

(1) numaralı eşitlikteki $P_{ik}^*(\theta)$ eğrilerine "işleyen karakteristik eğriler (operating characteristic curves) denilmektedir. ATM'de, her bir kategoriler arası eşik değeri için bir işleyen karakteristik eğri kestirilmelidir. Örneğin, beş kategoriye sahip bir madde için dört tane β_{ij} parametresi kestirilirken bir ayırt edicilik parametresi

(α_j) kestirilir. Kategori eşik parametreleri (β_{ij}), işleyen karakteristik eğrilerin konumunu belirler ve 0.50 olasılıkla j eşik kategorisinin üzerinde cevap vermek için gerekli yetenek düzeyini temsil eder. Madde ayırt edicilik parametresi ise bir maddenin farklı yetenek düzeyine sahip bireyleri ne derece iyi ayırabildiğinin bir ölçüsüdür. Genellikle α_j parametresinin değeri büyüdükçe daha dik, küçüldükçe daha dar işleyen karakteristik eğriler elde edilir. Çok kategorili MTK modellerinde ayırt edicilik parametresi dikkatli yorumlanmalı ve madde ayırt ediciliğinin büyüklüğü madde bilgi eğrileri ile birlikte değerlendirilmelidir (Embretson ve Reise, 2000; Tang, 2006).

MTK' da kullanılan bilgi terimi, bir maddenin kalitesinin istatistiksel bir göstergesidir. Madde cevapları θ -ölçeğinin bütün düzeylerinde bilgi sağlar, bilgi tek bir değerden ziyade kestirilen parametrenin bir fonksiyonudur. Madde bilgi fonksiyonu, bir test ya da maddeden elde edilen puanın ilgili örtük özellikteki farklılıkları ne derece ayırt ettiğinin bir göstergesidir (Reckase, 2009). Madde bilgi fonksiyonu her bir maddenin testin bütününün kesinliğine katkısını değerlendirmek için kullanılabilir, bu özelliğinden dolayı test geliştiricilerin madde seçimi için kullanışlı bir ölçüttür. MTK' da maddelerin birbirinden bağımsız olması sebebi ile her bir maddenin teste olan katkısı ayrı değerlendirilebilir. Böylece maddelerin test üzerindeki etkisi ve testin bireyin yeteneğini ölçmedeki gücü kolayca görülebilir. Bu durum test geliştirme sürecinde yarar sağlar (Hambleton ve Swaminathan, 1985). Bu çalışmada, PISA 2009 okuma becerileri tutum ölçeği maddelerinin değerlendirilmesinde madde parametrelerinin yanı sıra madde bilgi fonksiyonlarından ve test bilgi fonksiyonundan yararlanılmıştır.

Literatür incelendiğinde, daha çok ikili puanlanan maddeler için geliştirilen MTK modelleri üzerine araştırmalar yapıldığı görülmektedir. Bunun sebeplerinden biri çok kategorili MTK modelleri ile çalışmanın ikili puanlanan MTK modelleri ile çalışmaya nazaran daha karmaşık oluşudur (Maydeu-Olivares, Drasgow ve Mead, 1994). Ancak yine de kısmi puanlanan maddeler için ve Likert tipi ölçekler için çeşitli çalışmalar mevcuttur. Matteucci ve Straqualursi (2006), aşamalı tepki modeline göre öğrencilerin yazma testinden elde ettiği puanları değerlendirmiştir. Bir diğer çalışmada, MTK modellerinin ölçek geliştirme, değerlendirme ve düzenleme işlemlerinde uygulanması üzerinde durulmuş, aşamalı tepki modeli ile incelenmiştir. (Edelen ve Reeve, 2007). Bunların yanında kısmi puanlanan maddelere ilişkin yapılan çeşitli çalışmalar da bulunmaktadır. Bu araştırmanın amacı, ülkemizde uluslararası düzeyde öğrenci başarılarını belirlemede uygulanan PISA 2009 okumaya ilişkin tutum ölçeğinden elde edilen madde cevaplarını, çok kategorili sıralı puanlanan maddeler için geliştirilmiş olan aşamalı tepki modeli ile değerlendirmektir.

YÖNTEM

Araştırmanın Türü

Bu çalışmada, PISA 2009 kapsamında uygulanan okumaya yönelik tutum ölçeğinden elde edilen madde puanları aşamalı tepki modeli ile incelenmiştir. Var olan yöntem ve tekniklerin, gerçek veri üzerinde sınaması yapıldığından, betimsel araştırma niteliğindedir.

Evren ve Örneklem

Araştırmanın evrenini, Türkiye'de eğitim alan 15 yaş grubu öğrenciler oluşturmaktadır. Araştırmanın örneklemini ise Uluslararası Öğrenci Başarılarını Değerlendirme Programı (PISA)'nın uygulandığı 12 istatistikî bölge biriminden, 56 il ve okul türlerine göre tabakalandırılarak rastgele 160 okuldan seçilen 15 yaş grubu 4996 öğrenci oluşturmaktadır. Bu çalışma, PISA 2009 uygulamasının Türkiye örnekleminde okumaya yönelik tutum maddelerin tamamına yanıt veren 4198 öğrenci ile yürütülmüştür.

Veri Toplama Araçları

Çalışmada, öğrencilerin PISA 2009 değerlendirmesinde uygulanan öğrenci anketinde yer alan okumaya yönelik tutum maddelerine verdikleri yanıtlar veri olarak kullanılmıştır. Veriler, 2010 yılı Aralık ayında OECD resmi web sitesinde yayımlanmış olan rapordan alınmıştır. PISA öğrenci anketinde, okumaya yönelik tutuma ilişkin 11 madde bulunmaktadır. Öğrencilerden, her bir tutum maddesi için dört kategoriden birinde tepkide bulunmaları istenmiştir (1=hiç katılmıyorum, 2=katılmıyorum, 3=katılıyorum ve 4=tamamen katılıyorum). Maddelerdeki kategori sayıları eşittir. Olumsuz maddeler, "hiç katılmıyorum" kategorisi 4, "tamamen katılıyorum" kategorisi ise 1 olacak şekilde ters puanlanmıştır.

Veri Analizi

Araştırma verilerinin analizi iki aşamada gerçekleştirilmiştir. Veri analizinin ilk aşamasında, veri setinin MTK'nın tek boyutluluk ve yerel bağımsızlık varsayımlarını sağlayıp sağlamadığı incelenmiştir. Bunun için ilk olarak SPSS 20.0 programında 11 madde için açılımlı faktör analizi (AFA) yapılmıştır. AFA sonucu elde edilen özdeğeri 1'den büyük iki faktörden ilkinin tek başına toplam varyansın % 40'ını açıkladığı görülmüştür. Yamaç-eğitim grafiği ise baskın bir şekilde tek faktörde düşüş sergilemiştir. Döndürülmüş bileşenler matrisindeki faktör yükleri incelendiğinde (0.61-0.72) tutum anketini oluşturan maddelerin tek faktör altında toplanabileceği gözlenmiştir.

Anketin tek boyutlu yapıyı doğrulayıp doğrulamadığını incelemek amacıyla aynı zamanda LISREL 8.80 (Jöreskog ve Sörbom, 1999) programında doğrulayıcı faktör analizi (DFA) yapılmıştır. DFA sonuçlarına göre model-veri uyumu CFA, RMSEA ve NNFI değerleri kullanılarak değerlendirilmiştir. RMSEA, 0 ile 1 arasında değer almaktadır ve 0'a yaklaşması mükemmel uyuma işaret etmektedir (Bollen ve Curran, 2006; Tabachnick ve Fidell, 2007). Diğer uyum indekslerinden CFI ve NNFI'nin ise 0.95 ve üstü değerler alması model-veri uyumuna işaret etmektedir (Tabachnick ve Fidell, 2007; Raykov ve Marcoulides, 2006 v; Hu ve Bentler, 1995). Modele ait bu indeksler birlikte değerlendirildiğinde, tek faktör modeli kabul edilmiştir ($\chi^2_{(44)} = 1443.95$; RMSEA=0.087; CFA=0.96 ve NNFI=0.95). AFA ve DFA sonucu elde edilen tüm bulgulara dayalı olarak ölçeğin tek boyutlu bir yapıya sahip olduğu söylenebilir.

MTK'nın bir diğer varsayımı yerel bağımsızlıktır. Yerel bağımsızlık, bir bireyin bir testteki farklı maddelere verdiği cevapların birbirinden istatistiksel olarak bağımsız olması demektir. Bu varsayımın doğru olması için bir bireyin, bir maddedeki performansının diğer maddelerdeki performansını etkilememesi gerekir. Bu varsayımın ihlali aynı zamanda tek boyutluluk varsayımının da ihlali demektir. Yerel bağımsızlık, maddeler arası ilişkilerin açıklanması için sadece bir yeteneğin yeterli olması demektir (Hambleton ve Swaminathan, 1985). Dolayısıyla tek boyutluluk varsayımını sağlayan tutum ölçeğindeki 11 maddenin yerel bağımsızlık varsayımını da sağladığı kabul edilmiştir.

Veri analizinin ikinci aşamasında, MULTILOG (Thissen, 1991) programında aşamalı tepki modeli (graded response model) seçilerek madde kalibrasyonu yapılmıştır. Modele ait madde parametreleri Marjinal Maximum Likelihood (MML) yöntemi ile kestirilmiştir.

BULGULAR

Bu bölümde araştırmanın amacına göre elde edilen bulgulara yer verilmektedir.

PISA 2009 okumaya yönelik tutum anketinden elde edilen verilerin aşamalı tepki modeli ile kalibre edilmesi sonucu negatif log likelihood (-2*LL) değeri 24385.1 olarak bulunmuştur. Maksimum likelihood kestiriminde negatif log likelihood değeri, verinin modelden uzaklaşma derecesini göstermektedir (Embretson ve Reise, 2000). Ankete ait marjinal güvenilirlik katsayısı 0.87 olarak elde edilmiştir. Marjinal güvenilirlik, tüm yetenek düzeyindeki öğrencilerden tahmin edilen koşullu standart hatalarının ortalaması olarak ele alınan toplam güvenilirliği temsil etmektedir (DCAS 2010–2011, *Tecnical Report*).

Maddelerin veriye uyum düzeyleri, gözlenen ve beklenen oranlar arasındaki fark aracılığı ile incelenebilir. Gözlenen ve beklenen oranlar arasındaki farklar "artık" olarak da isimlendirilir. Embretson ve Reise (2000) , artıkların sifıra yaklaşmasının (<0.01) modelin veri ile uyum gösterdiğinin bir ölçütü olarak alınabileceğini belirtmiştir. Tablo 1'de, kategorilere yanıt veren öğrenci sayıları ile gözlenen-beklenen frekans oranları ve bu frekans oranları arasındaki fark (artık) değerleri verilmiştir. Tablo 1'de yer alan 11 maddenin her bir kategorisine ait beklenen ve gözlenen oranları arasındaki farklar incelendiğinde bütün artıkların .01 değerinden daha küçük olduğu görülmektedir. Bu bulguya dayalı olarak ATM modelinin veriye uyum gösterdiği söylenebilir.

Tablo 1: Kategorilere Ait Gözlenen ve Beklenen Frekans Oranları

Maddeler		Kategoriler			
		1	2	3	4
1	f	141	716	1472	1869
	g	0.0336	0.1706	0.3506	0.4452
	b	0.0352	0.1745	0.3489	0.4413
	fark	0.0016	0.0039	0.0017	0.0039
2	f	355	1096	2029	718
	g	0.0846	0.2611	0.4833	0.1710
	b	0.0903	0.2595	0.4801	0.1700
	fark	0,0057	0,0016	0,0032	0,0010
3	f	391	976	2058	773
	g	0,093	0,233	0,490	0,184
	b	0,100	0,231	0,489	0,180
	fark	-0,007	0,001	0,002	0,004
4	f	277	1090	1794	1037
	g	0,066	0,260	0,427	0,247
	b	0,070	0,263	0,424	0,243
	fark	-0,004	-0,004	0,004	0,004
5	f	257	688	2204	1049
	g	0,061	0,164	0,525	0,250
	b	0,065	0,169	0,520	0,246
	fark	-0,004	-0,005	0,005	0,004
6	f	102	191	1401	2504
	g	0,024	0,046	0,334	0,597
	b	0,027	0,048	0,331	0,594
	fark	-0,003	-0,003	0,003	0,002
7	f	247	802	2219	930
	g	0,059	0,191	0,529	0,222
	b	0,063	0,195	0,524	0,218
	fark	-0,004	-0,004	0,005	0,004
8	f	309	989	1860	1040
	g	0,074	0,236	0,443	0,248
	b	0,077	0,236	0,442	0,245
	fark	-0,003	-0,001	0,001	0,003
9	f	288	959	1695	1256
	g	0,069	0,228	0,404	0,299
	b	0,071	0,229	0,400	0,299
	fark	-0,003	-0,001	0,004	0,000
10	f	169	430	2190	1409
	g	0,040	0,102	0,522	0,336
	b	0,044	0,108	0,517	0,331
	fark	-0,004	-0,006	0,005	0,005
11	f	285	650	1980	1283
	g	0,068	0,155	0,472	0,306
	b	0,072	0,157	0,467	0,305
	fark	-0,004	-0,002	0,005	0,001

f: frekans, g: gözlenen, b: beklenen

Elde edilen bir diğer bulgu, Tablo 2’de de görülmekte olan ATM’ ye dayalı olarak kestirilen madde parametreleridir.

Tablo 2: Aşamalı Modele Göre Kestirilen Parametre Değerleri

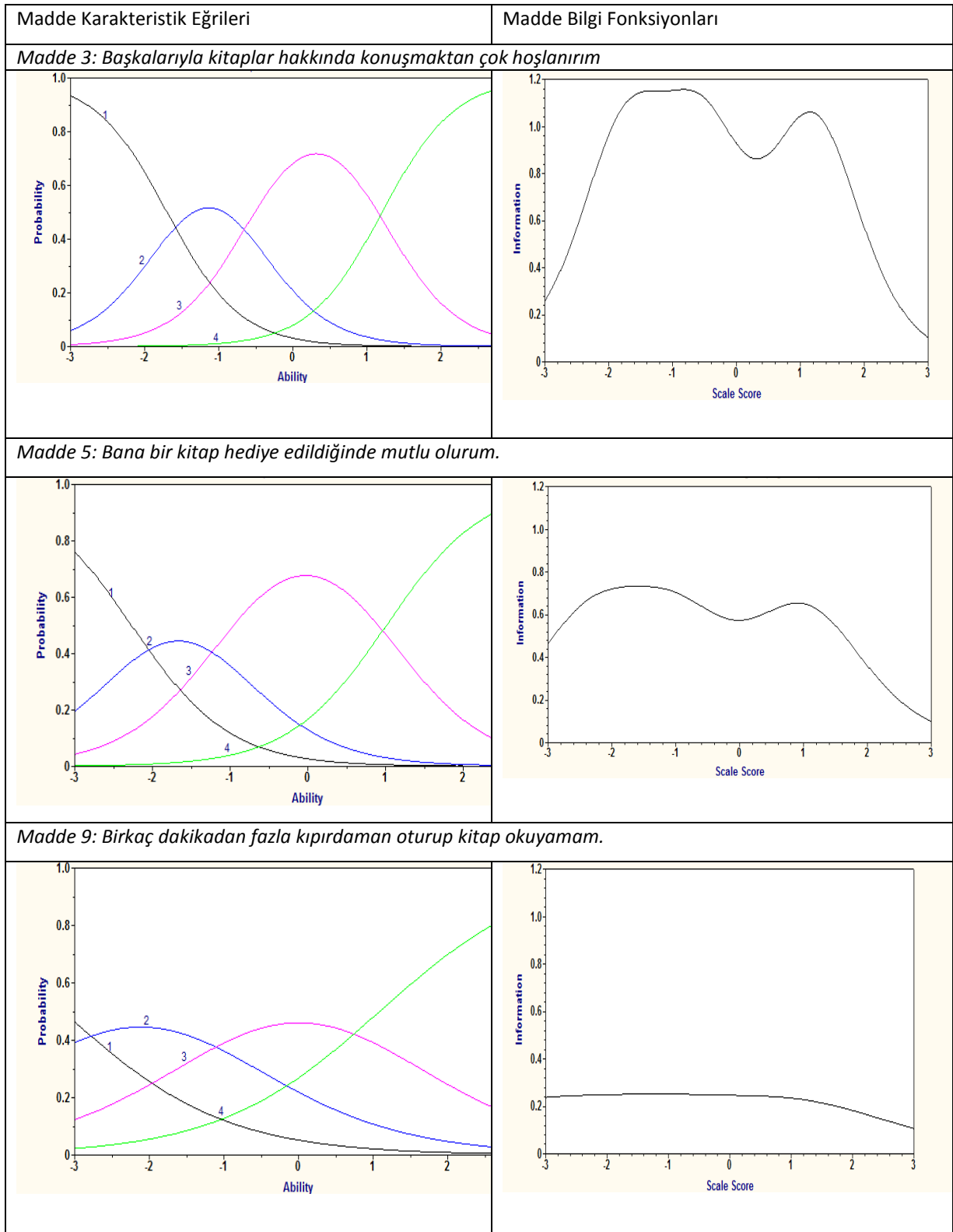
Madde	α_i (SE)	β_1 (SE)	β_2 (SE)	β_3 (SE)
1	1.35 (0.05)	-3.03 (0.13)	-1.30 (0.05)	0.24 (0.04)
2	1.85 (0.06)	-1.85 (0.06)	-0.53 (0.03)	1.31 (0.04)
3	2.03 (0.06)	-1.70 (0.05)	-0.57 (0.03)	1.20 (0.04)
4	1.60 (0.05)	-2.20 (0.08)	-0.63 (0.04)	1.02 (0.04)
5	1.59 (0.05)	-2.27 (0.08)	-1.07 (0.04)	1.01 (0.04)
6	1.76 (0.07)	-2.79 (0.12)	-2.05 (0.07)	-0.33 (0.03)
7	1.94 (0.06)	-2.08 (0.07)	-0.87 (0.03)	1.04 (0.04)
8	1.59 (0.05)	-2.14 (0.07)	-0.71 (0.04)	1.02 (0.04)
9	0.92 (0.04)	-3.16 (0.16)	-1.08 (0.07)	1.08 (0.07)
10	1.32 (0.05)	-2.88 (0.12)	-1.68 (0.06)	0.71 (0.04)
11	1.66 (0.05)	-2.14 (0.08)	-1.07 (0.04)	0.73 (0.04)

Tablo 2 incelendiğinde, eğitim parametrelerinin 0.92 ile 2.03 arasında değerler aldığı görülmektedir. Eğitim parametresi, madde ile ölçülmek istenen yetenek arasındaki güçlü ilişkiyi yansıtır. Bu bağlamda pozitif ve yüksek eğitim değerleri tercih edilir. Baker (2001, s.35), maddelerin ayırt edicilik parametre değerlerine göre çok düşük (0.01-0.34), düşük (0.35-0.64), orta (0.65-1.34), yüksek (1.35-1.69) ve çok yüksek (>1.70) ayırtıcılığa sahip maddeler olarak sınıflandırılabilceğini belirtmiştir. Tablo 2 incelendiğinde 1, 9 ve 10. maddeler orta düzeyde ayırt ediciliğe sahip iken geriye kalan 8 maddenin yüksek ve çok yüksek ayırt ediciliğe sahip olduğu söylenebilir. Tüm maddeler içerisinde 3. maddenin “Başkalarıyla kitaplar hakkında konuşmaktan çok hoşlanırım” en yüksek ($\alpha_i=2.03$), 9. maddenin “Birkaç dakikadan fazla kıpırdamadan oturup kitap okuyamam” ise en düşük ($\alpha_i=0.92$) eğitim parametresine sahip olduğu görülmektedir.

Tablo 2’de ayrıca kategoriler arası eşik değerleri (β_i) yer almaktadır. Bu değerler, ATM’de kategoriler arasında sıralıdır. β_i , öğrencilerin 0.50 olasılıkla eşik değerinin üzerinde cevap vermek için gerekli olan tutum düzeyini gösterir (Matteucci ve Stracqualursi, 2006). Buna göre bir öğrencinin ilk kategoriyi cevaplaması için düşük yetenek ya da tutuma sahip olması yeterli olmakta iken, bir üst kategori için daha yüksek tutum gerekmektedir. Kategori arası eşik parametreleri, -3.16 ile 1.31 değerleri arasında değişmektedir. Eşik parametre değerlerinin çoğunun negatif değerler almasına dayalı olarak kategorilerdeki cevapların daha çok tutumun düşük düzeyi ($\theta < 0$) tarafından desteklendiği söylenebilir.

Tablo 2’de verilen bilgilere göre, 9. maddede birinci kategoriden ikinci kategoriye geçiş için eşik parametresi, -3 yetenek düzeyinin altında (-3.16) kestirilmiştir. Bu durum, ilk kategoriyi az kişinin yanıtlamasından ve maddenin ölçmek istediği özellikle ilişkili olmamasından kaynaklanmış olabilir (Embretson ve Reise). Nitekim Tablo 1 incelendiğinde, 9. maddenin birinci kategorisine cevap veren öğrenci sayısının (288) diğer kategorilere cevap veren öğrenci sayısından oldukça az olduğu görülmektedir, öğrencilerin sadece % 7 ‘si ilk kategoriyi seçmiştir. Tablo 2’de verilen 9. maddeye ait ayırt edicilik parametresinin düşük olması (0.92) da bu durumun oluşmasının bir diğer sebebi olarak gösterilebilir.

Maddelerin istatistiksel özelliklerine dair bilgi edinmenin bir diğer yolu madde karakteristik eğrilerini ve bilgi fonksiyonu grafiklerini incelemektir. Şekil 1’de üç maddeye ait kategori karakteristik eğrileri verilmiştir. Bu eğriler, her bir yanıtlama düzeyinin tek başına yeteneğin bir fonksiyonu olduğunu göstermektedir. Bu maddelerden, 3. madde en yüksek eğime, dolayısı ile ayırtıcılığa ($\alpha_i=2.03$); 5. madde orta düzeyde bir eğime ($\alpha_i=1.59$) ve 9. madde en düşük eğime sahip olması ($\alpha_i=0.92$) nedeni ile seçilmiş ve örnek teşkil etmesi açısından grafikleri verilmiştir.

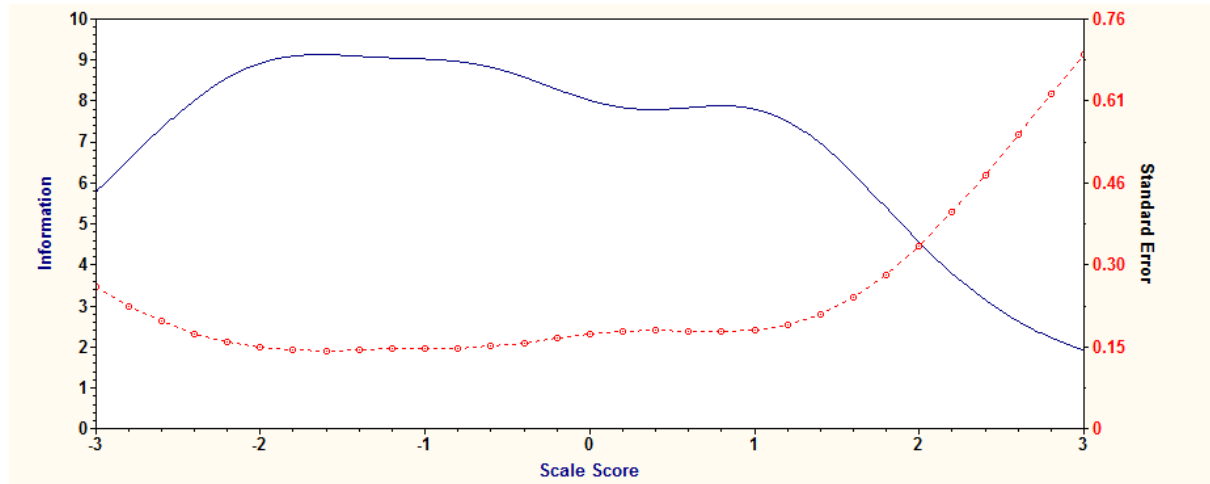


Şekil 1: PISA tutum anketinde yer alan maddelerin farklı eğitim ve kategori eşik parametrelerine göre karakteristik eğrileri

Şekil 1’de yer alan madde yanıtları 1 (hiç katılmıyorum) ve 4 (tamamen katılıyorum) arasında değerler almakta, 1 en düşük tutumu, 4 ise en yüksek tutumu göstermektedir. Buna göre, 3. maddeye ait madde karakteristik eğrisinde 4. kategori monoton artan bir eğim sergilemektedir. Başka deyişle, yetenek düzeyi ya da tutum düzeyi arttıkça 4. kategorinin seçilme olasılığı artmaktadır. Bunun aksine 1. kategori monoton azalan bir eğime sahiptir. Bu durum, yetenek düzeyi arttıkça o kategorinin seçilme olasılığının düşük olduğunu göstermektedir. Geriye kalan 2. ve 3. kategoriler ise düşük ve orta düzeydeki yeteneklerde artan bir eğim, yüksek yetenek düzeylerinde azalan bir eğim sergilemektedir. Düşük eğim parametresine sahip olan 9. maddeye ait kategori karakteristik eğrisi incelendiğinde, kategorilerin çok geniş bir yetenek alanında yanıtlanma olasılığına sahip olduğu görülebilir. Orta düzeyde eğim parametre değerine sahip olan 5. maddenin ise 3. maddeden daha geniş bir alanda yanıtlanma olasılığı bulunduğu görülmektedir.

Şekil-1’deki madde bilgi fonksiyonları, madde karakteristik eğrilerindeki farklılıklar sonucu o maddenin ölçmenin kesinliğine olan etkisini de göstermektedir. En yüksek eğim parametre değerine sahip olan 3. madde aynı zamanda en çok bilgiyi veren maddedir. Buna karşılık 5. ve 9. maddeler, en çok bilgiyi düşük tutum düzeyinde vermektedirler.

Madde bilgi fonksiyonlarının yanı sıra Şekil 2’de görülen tüm maddelerden elde edilen toplam bilginin yer aldığı test bilgi fonksiyonu incelenmiştir.



Şekil 2: Aşamalı tepki modeline göre elde edilen test bilgi fonksiyonu

Şekil 2, testteki bilginin en çok -2 ile -1 yetenek ya da tutum düzeyleri arasında elde edildiğini göstermektedir. Madde kalibrasyon sonuçları incelendiğinde testin fazla bilgiyi -1.6 yetenek düzeyinde (bilgi değeri: 9.118) verdiği görülmüştür. Testin verdiği bilgi özellikle 1.8 yetenek düzeyinden sonra azalmıştır. Bu durum standart hatanın da artmasına sebep olmuştur.

TARTIŞMA VE SONUÇ

Öğrencilerin herhangi bir konu alanındaki başarılarının o alana karşı tutumları ile ilişkili olduğu düşünülür. Bu nedenle öğrencilerin tutumlarını belirlemeye yönelik çalışmaların önemi göz ardı edilemez. Yapılan çalışmaların nitelikli olması açısından uygulanan ölçme araçlarının da nitelikli olması gerekir. Bu nedenle ilgili çalışmada, PISA 2009 okumaya yönelik tutum anketinin maddeleri çok kategorili madde tepki modellerinden Likert tipi ölçekler için uygun olan “aşamalı tepki modeli” ne göre incelenmiştir. Yapılan açımlayıcı ve doğrulayıcı faktör analizi sonuçlarına göre anketin tek boyutlu bir yapıya sahip olduğu görülmüş, tek boyutluluk varsayımını sağlayan maddelerin aynı zamanda yerel bağımsızlık varsayımını sağladığı kabul edilmiştir. Gözlenen ve beklenen frekans oranları arasındaki farkın küçük olması aşamalı tepki modelin veriye uygun olduğunu, dolayısı

ile PISA 2009 okuma yönelik tutum maddelerinin seçiminde ve değerlendirmesinde uygun bir model olabileceğini göstermiştir.

Aşamalı tepki modeli, maddelerin istatistiksel özellikleri hakkında derinlemesine bilgi sahibi olunmasını sağlamıştır. Bu model, klasik test kuramından (KTK) farklı olarak ayırt edicilik parametrelerinin yanında madde karakteristik eğrilerinin de birlikte yorumlanmasına imkan vermiştir. Bu durum MTK'nın bilgi kuramına dayanmasından ve bilgiyi maksimize eden "En Çok Olabilirlik" (Likelihood) kestirim yöntemini kullanmasından kaynaklanmaktadır (Yurdugül, 2003). Açıklayıcı ve doğrulayıcı faktör analizinden elde edilen bulgular, KTK kapsamında maddelerin birbiriyle ve ölçmek istediği özelliklerle ilişkilerinin yüksek düzeyde olduğunu gösteren bir sonuç olarak yorumlanabilir. KTK kapsamında elde edilen bu bilginin ötesinde, MTK yetenek ya da tutum düzeyine göre öğrencileri birbirinden ayırmış, ölçmek istenen özellik hakkında en fazla bilgiyi veren tutum düzeyinin belirlenmesini sağlamıştır. Elde edilen bulgulara göre; okumaya yönelik tutumları açısından öğrencileri en iyi ayırt eden dolayısıyla en çok bilgi veren maddenin 3. madde, en kötü ayırt eden ve en az bilgiyi veren maddenin ise 9. madde olduğu sonucuna ulaşılmıştır. Matteucci ve Stracqualursi (2006), eğitim parametresinin madde ile ölçmeye çalıştığı örtük özellik arasındaki gücün bir göstergesi olduğunu belirtmiştir. Buna göre, 3. maddenin okumaya yönelik tutum ile en güçlü ilişkiye sahip, 9. maddenin ise en zayıf ilişkiye sahip madde olduğu söylenebilir.

Kategoriler arası eşik parametreleri, birinci kategoriden ikinciye ve ikinci kategoriden üçüncüye geçişte daha çok düşük tutum düzeyi tarafından desteklenirken, üçüncü kategoriden dördüncü kategoriye geçiş, yüksek düzeyde tutum gerektirmiştir. Bazı maddelerde dördüncü kategoriye geçişte, tutum düzeyi düşük olarak bulunmuştur. Özellikle 1. ve 6. maddelerde dördüncü kategoriye geçiş için, üçüncü kategori eşik parametre değeri çok küçük olarak elde edildiğinden, öğrencilerin düşük tutum düzeyine sahip olması yeterli görülmüştür. Bunların yanında 10 ve 11. maddelerde de dördüncü kategoriye geçiş, tutumu düşük öğrenciler tarafından desteklenmiştir.

Araştırma bulgularına dayalı olarak, PISA 2009 okumaya yönelik tutum anketini oluşturan maddelerin (8 maddenin) büyük çoğunluğunun ayırt ediciliğinin yüksek olduğu sonucuna ulaşılmıştır. PISA, öğrencilerin performansını değerlendiren ve öğrencilerin performanslarındaki farklılıkları açıklamaya yardımcı olabilecek öğrenci, aile ve öğretim faktörlerine ilişkin bilgi toplayan en kapsamlı uluslararası eğitim araştırmalarından biridir. Bu bağlamda geliştirilen ölçme araçlarının tüm ülkelerde geçerli ve güvenilir olmasını sağlamak, kültürel ve dilsel farklılıkları en aza indirmek amacıyla oldukça yoğun bir çaba ve kaynak sarf edilmektedir. PISA'dan elde edilen bulguların yüksek güvenilirlik ve geçerliğe sahip olduğu iddia edilmektedir (MEB, 2010). Bu araştırmadan elde edilen bulgular da PISA'nın bu iddiasını destekler niteliktedir.

Not: Bu çalışma 25-27 Nisan 2013 tarihlerinde Antalya'da 28 Ülkenin katılımıyla düzenlenen " International Conference on New Trends in Education – ICONTE – 2013 "da sözlü bildiri olarak sunulmuştur.

KAYNAKÇA

Doğan, N. ve Tezbaşaran, A.A. (2003). Klasik test kuramı ve örtük özellikler kuramının örneklem bağlamında karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 25, 58-67.

DCAS (2010–2011). Delaware Comprehensive Assessment System. Evidence of Reliability and Validity. *Volume 4, American Institutes for Research..*

Baker, F. B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation. (ERIC Document Reproduction Service No. ED458219).

Bollen, K. A. & Curran, P. J. (2006). *Latent Curve Models: A Structural Equation Perspective*. New Jersey: John Wiley & Sons, Inc.

Edelen, M. O. & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res*, 16(5), 5-18.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ:Lawrence Erlbaum Associates, Inc.

Hambleton, R.K. ve Jones, R.W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*.12(3), 38-47.

Hambleton, R. K & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Norwell, MA:kluwerAcademic Publishers.

Jöreskog, K. G. ve Sörbom, D. (1999). *LISREL 8: Structural Equation Modeling with the Simplis Command Language*. USA: Scientific Software international, Inc.

Maydeu-Olivares, A., Drasgow, F. & Mead, A. D. (1996). Distinguishing among parametric item response models for polytomous ordered data. *Applied Psychological Measurement*, 18(3), 245-256.

Milli Eğitim Bakanlığı. (2010). *PISA 2009 projesi ulusal ön raporu*. Ankara: Milli Eğitim Basımevi.

Matteucci, M. & Stracqualursi, L. (2006). Student assessment via graded response model. *Statistica*, 4, 435-447.

Raykov, T. & Marcoulides, G. A. (2006). *A First Course in Structural Equation Modeling (2nd Edt)*. London: Lawrence Erlbaum Association.

Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer.

Tabachnick, B. G. & Fidell, L.S. (2007). *Using Multivariate Statistics (5th Edt)*. USA: Pearson.

Tang, K. L. (2006). *Polytomous item response theory models and their applications in large-scale testing programs: Review of literatüre*. Educational Testing Service: Princeton, New Jersey.

Thissen, D. (1991). *MULTILOG user's guide (Version 6.0)*. Mooresville, IN: Scientific Software.

Yurdugül, H. (2003). Orta Öğretim Kurumları Seçme ve Yerleştirme Sınavının Madde Yanlılığı Açısından İncelenmesi. *Hacettepe Üniversitesi (Yayımlanmamış DoktoraTezi)*.