



Üstbilişsel İzleme ve Matematiksel Yeterlilikler: Bilişsel Tanı Modeli ve Sinyal Tespit Kuramı Yaklaşımı *

Oğuz Tahsin Başokçu ¹, Mehmet Akif Güzel ²

Öz

Öğrencilerin matematik becerilerinin belirlenmesinde, diğer birçok sınıf içi değerlendirme yanında, PISA (Uluslararası Değerlendirme Programı) ve TIMMS (Uluslararası Matematik ve Fen Eğilimleri Araştırması) gibi çeşitli ülkelerde standart olarak kullanılan ölçme araçları bulunmaktadır. Ülkelerarası uygulanan bu geniş ölçekli testlerdeki soru içerikleri, soru yanıtlama türü, puanlanmalar ve analiz yöntemleri yıllar içinde çeşitlenmiştir. Bu araştırmada, PISA matematik testine ait becerilerin değerlendirmesinde daha önceden kullanılmamış olan Bilişsel Tanı Modeli (BTM) ve Sinyal Tespit Kuramı'na (STK) ait analiz yöntemleri birlikte kullanılarak bu becerilerin tek bir test prosedürü içinde belirlenmesi amaçlanmıştır. Bu amaç doğrultusunda rastgele örnekleme yöntemiyle İzmir ilinden belirlen 6. Sınıf öğrencilerine ($N=230$), PISA matematik testi eşdeğerliğinde farklı madde formatlarının kullanıldığı (çoktan seçmeli, doğru yanlış, kısa cevaplı) ve doğru/yanlış olarak dikatomik şekilde puanlanan 12 sorudan oluşan bir matematik testi uygulanmıştır. BTM ölçümleri DINA (the Deterministic Input Noisy Output and Gate) model kullanılarak gerçekleştirilmiştir. STK yöntemiyle ise, bireyin doğru ve yanlış yanıtlarını ayırt edebilme becerisi olarak tanımlanan "üstbilişsel izleme" performansının ölçülmesi amacıyla, katılımcılardan her bir soru maddesi için sırasıyla, soruyu çözüp çözemeyeceğini belirtmesi, doğru yanıtlayabileceğini düşünmemiş olsa bile bir tahminde bulunması ve verdiği yanıtların doğruluğundan ne kadar emin olduklarını derecelendirmeleri istenmiştir. Üstbilişsel izleme performansı, özetle, katılımcıların gerçekte doğru olan yanıtlarında soruyu yanıtlayabileceğini seçip yüksek eminlik düzeyleri vermesi, yanlış olan yanıtlarında ise soruyu yanıtlayabileceğini düşünmemesini ve tahminleri için düşük eminlik düzeyleri vermesiyle, yanıtlarının doğru ve yanlış olarak ayırımını ne kadar iyi yapabildiğinin puanlanmasıyla elde edilmiştir. Sonuçlar, PISA testinin BTM yöntemine iyi uyum sağladığını göstermiş ve BTM yöntemiyle belirlenen dört temel

Anahtar Kelimeler

Bilişsel Tanı Modeli
Sinyal Tespit Kuramı
Üstbiliş
Üstbilişsel İzleme
Matematik
PISA testi

Makale Hakkında

Gönderim Tarihi: 17.06.2018
Kabul Tarihi: 02.06.2020
Elektronik Yayın Tarihi: 12.11.2020

DOI: 10.15390/EB.2020.7991

* Bu makale 17. EARLI (The European Association for Research on Learning and Instruction) konferansında sözlü bildiri olarak sunulmuştur.

¹ Ege Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Türkiye, oguzbasokcu@gmail.com

² Abdullah Gül Üniversitesi, İnsan ve Toplum Bilimleri Fakültesi, Psikoloji Bölümü, Türkiye, akif.guzel@agu.edu.tr

beceriden (“ilişkilendirme ve iletişim”, “matematikleştirme”, “akıl yürütme ve strateji geliştirme” ve “sembolik ve teknik dil kullanımı”) özellikle “akıl yürütme ve strateji geliştirme” becerisi yüksek olan öğrencilerin üstbilişsel izleme performanslarının da yüksek olduğu görülmüştür. Araştırma, doğru ve yanlış yanıtları yüksek oranda ayırt edebilme becerisinin gerçekte hangi özelliklerle ilişkili olduğunun bulunması konusundaki çalışmalarına katkı sağlamaktadır. Bulgular temelinde, üstbilişsel izleme becerisinin özgül olarak “akıl yürütme ve strateji geliştirme” olarak gözlenen alt özelliğe sahip olmakla ilişkili olabileceği önerilmiştir. Ek olarak, PISA testinde -veya herhangi bir başka de olabileceği gibi- üstbilişsel izleme performansının STK ölçüm metodu gibi görece kolay ek bir test prosedürüyle ölçülerek, testte ölçülen becerilerin yanında öğrencilerin diğer ilgili üst-düzyer becerileri hakkında da bilgiler elde edilebileceği önerilmektedir.

Giriş

Eğitimde kullanılan ölçme yöntemlerinin birincil amacı, öğrencilerin sahip olduğu akademik performansın düzeylerini en doğru ve en kesin şekilde belirleyebilmektir. Bu amaçla, eğitim alanında kullanılan ölçme yöntemlerinde, örneğin, bilişsel beceriler ve akademik başarının belirlenmesinde çok sayıda ölçme stratejisi ve yaklaşımı bulunmaktadır (ör., Bean ve Peterson, 1998; Wragg, 2001; Lindblom-Ylänne, Pihlajamäki ve Kotkas, 2006). Eğitim sürecinin önemli bir ögesi olan performans değerlendirmesi için çeşitli değerlendirme araçlarının sınıf içi değerlendirmelerde sıklıkla kullanılmasının yanında, çeşitli standart testler de kullanılmaktadır. Örneğin, öğrencilerin matematik, fen bilgisi, dil becerisi gibi temel alanlardaki akademik performanslarının ülkeler arasında ne düzeyde farklılaştığını belirlemek ve bu sonuçları göz önünde bulundurarak eğitim politikalarına yön verebilmek amacıyla da dünyada yaygın olarak kullanılan bir takım standart testler bulunmaktadır. İlk olarak 2000 yılında kullanılmaya başlanan ve yıllar içinde hem içerik hem de puanlama bakımından geliştirilmeye devam edilen PISA (Uluslararası Öğrenci Değerlendirme Programı) ve TIMSS (Uluslararası Matematik ve Fen Eğilimleri Araştırması) sınavları bu tür değerlendirme yöntemlerine örnek olarak verilebilir.

Özgül olarak PISA testinde ölçülen matematik becerilerine odaklanan bu çalışmada, 6. sınıf öğrencilerinin PISA matematik testinde gösterdikleri yeterliklere ne düzeyde sahip olduklarının, daha önce PISA testi değerlendirmesinde kullanılmamış olan iki yöntemin aynı anda kullanılarak belirlenmesi amaçlanmıştır. Bu yöntemlerden ilki Bilişsel Tanı Modeli ([BTM]; Cognitive Diagnostic Model), bir diğeri ise Tip-2 Sinyal Tespit Kuramı'na ([STK]; Signal Detection Theory) ait ölçme yöntemidir. Araştırmada, öğrencilerin matematik becerilerinin BTM'ye göre belirlenmesi ve testte ölçülen matematik alanına ait bilişsel yeterlikler ile kısaca 'bireyin doğru ve yanlış yanıtlarını ne kadar iyi şekilde ayırt edebildiği'ne karşılık gelen “üstbilişsel izleme” (metacognitive monitoring) (ör., Higham, 2002; Higham ve Gerrard, 2005; Güzel ve Higham, 2013) performansı arasındaki olası ilişkilerin belirlenmesi amaçlanmıştır (bkz. Karakelle ve Saraç, 2010)¹. Özetle, BTM ve Tip-2 STK metoduna ait hesaplamalar sonucunda, üstbilişsel izleme becerisinin, PISA testinde standart olarak ölçülen matematik becerisi için ortaya çıkarılması beklenen örtük sınıfların hangisi veya hangileri ile ilişkili olduğunun tek bir test uygulamasıyla belirlenmesi amaçlanmıştır. Alanyazındaki çalışmalar incelendiğinde, ilerleyen bölümlerde detaylandırılan ve özellikle Tip-2 STK'ya ait ölçme yöntemlerinin, öğrencilerin PISA testindeki yanıtlarının doğruluklarını ne düzeyde ayırt edebildiklerini (diğer bir ifadeyle, üstbilişsel izleme performanslarını) doğrudan incelememeleri nedeniyle, yürütülen araştırmanın alanyazına özgün bir katkı sağlayabileceği düşünülmektedir.

¹ Çalışma, özellikle Tip-2 Sinyal Tespit Kuramı'na ait varsayımları temel alıp bu kurama ait ölçme yöntemini kullandığı için, bu alanyazına ait olan ancak, örneğin eğitim alanında ‘üstbilişsel kalibrasyon (metacognitive calibration)’ ya da ‘üstbilişsel kesinlik (metacognitive accuracy)’ olarak kullanılan terimler için (bkz., Pieschel, 2009), makelede “üstbilişsel izleme (metacognitive monitoring)” terimi kullanılmıştır (ör., Higham, 2011).

Devam eden bölümlerde, sırasıyla PISA matematik testinde ölçülen yeterlikler açıklandıktan sonra, BTM ve STK'nın bir türü olan Tip-2 ölçme yöntemleri detaylandırılmaktadır.

PISA Matematiksel Yeterlikler

Matematikte tanımlama, anlama ve bireyin şimdiki ve gelecekteki özel, iş ve sosyal hayatında matematiğin oynadığı rol hakkında, yaşitlarına göre sağlam şekilde temellendirilmiş kararlar verebilme becerisine karşılık gelen “matematik okuryazarlığı” (matematics literacy) (OECD, 2003) konusunda kazandırılması amaçlanan becerilere yönelik olarak çeşitli yaklaşımlar bulunmaktadır (bkz., Stacey ve Turner, 2014). Ancak, özellikle PISA ve TIMSS araştırmaları, matematik yeterlikleri konusunda, matematiğe ait olası tüm ana becerileri kapsayan sorularla daha bütünleştirici ve ölçüm temelli tanımlamalar kullanılmaktadır (Albacete vd., 2016; Gierl, Alves ve Majeau, 2010). Araştırmanın odağında bulunan PISA testi, sanayileşmiş ülkelerin hükümetler arası bir organizasyonu olan Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) tarafından gerçekleştirilmiş ve ilk olarak 2000 yılında uygulanan, 15 yaşındaki öğrencilerin matematik okuryazarlığı, fen okuryazarlığı ve okuma becerilerini 3 yılda bir ölçen uluslararası bir değerlendirme sistemidir (OECD, 2010). Yedi temel başlıkta ifade edilen bu yeterlikler, Tablo 1'deki gibi tanımlanmıştır (OECD, 2019).

Tablo 1. PISA Matematik Testinde Ölçülen Yeterlikler

İletişim kurma:	<ul style="list-style-type: none"> – Durumun zihinsel modelini oluşturmak için ifadeleri, soruları, görevleri, nesne veya imgeleri okumak, çözmek ve anlamlandırmak – Bir çözümü ifade etmek, bir sonuca ulaşmakta kullanılan işlemi göstermek ve/veya ara matematiksel sonuçları özetlemek ve göstermek – Problem bağlamındaki açıklamaları ve iddiaları oluşturmak ve bunları ifade etmek
Matematikleştirme:	<ul style="list-style-type: none"> – Gerçek hayatta dair problemlerde altta yatan matematiksel değişkenleri ve yapıları belirlemek ve bunların kullanılabilmesi için varsayımlarda bulunmak – Bağlam farkındalığını kullanarak matematiksel çözüm süreçlerini yönlendirmek ve bunları hızlandırmak; örneğin, bağlama uygun kesinlik düzeyi üzerinde çalışmak gibi – Kullanılan bir matematiksel modelin sonucu olan bir matematiksel çözümün büyüklüğünü ve sınırlarını anlamak
Temsil etme:	<ul style="list-style-type: none"> – Gerçek hayata dair bilginin matematiksel temsilini oluşturmak – Problemlerle ilgilenirken çeşitli temsilleri anlamlandırmak, bunlar arasında bağlantılar kurmak ve bunları kullanmak – Bir durum ya da kullanımla ilgili olan çeşitli formatlardaki matematiksel çıktıları yorumlamak; iki ya da daha fazla temsili bir durumla ilintili olarak karşılaştırmak veya değerlendirmek
Akıl yürütme ve kanıt gösterme:	<ul style="list-style-type: none"> – Belirlenmiş ya da oluşturulmuş gerçek hayata dair bir durumun temsiliğini açıklamak, savunmak veya haklı çıkarmak – Matematiksel bir sonuç ya da çözümün belirlenmesinde kullanılan sonuca dair süreçleri ve işlemleri açıklamak, savunmak veya haklı çıkarmak – Matematiksel bir çözüme ulaşmadaki bilgi parçalarını birleştirmek, genellemeler yapmak veya çok-adımlı savlar oluşturmak – Matematiksel çözümler üzerinde düşünmek ve matematiksel bir çözüme destekleyen, çürüten veya bunu bağlamsal içinde oluşturulan bir problemde değerlendiren açıklamalar veya savlar oluşturmak

Tablo 1. Devamı

Problem çözmede strateji geliştirme:	<ul style="list-style-type: none"> – Bağlam içinde kullanılan (oluşturulan) problemleri matematiksel olarak farklı şekilde ifade etmek için plan veya strateji seçmek veya oluşturmak – Matematiksel bir çözüme, sonuca veya genellemeye giden çok-adımlı bir süreç boyunca etkin ve kesintisiz kontrol mekanizmalarını çalıştırmak – Matematiksel bir sonuca ulaşmak için bir strateji oluşturup bunu bağlam içinde oluşturulan bir probleme dair matematiksel bir çözümü yorumlamak, doğrulamak ve değerlendirmek için uygulamak
Sembolik, formel ve teknik dil ve işlemleri kullanma:	<ul style="list-style-type: none"> – Sembolik ve formel bir dille, gerçek hayattan bir problemi temsil etmek için uygun değişkenler, semboller, diyagramlar ve standart modelleri kullanmak – Tanımlar, kurallar ve formel sistemler ve de algoritmaların kullanımına dayan formel yapıları anlamak ve kullanmak – Problemin bağlamıyla matematiksel çözümün temsili arasındaki bağlantıyı kavramak – Bu kavrayışı, bağlam içindeki çözümü yorumlamaya yardımcı olmak üzere kullanmak ve çözümün elverişli olup olmadığını ve olası çözümlerin sınırlıklarını tartmak
Matematiksel araçları kullanma:	<ul style="list-style-type: none"> – Matematiksel yapıları fark etmek veya matematiksel ilişkileri tasvir edebilmek için matematiksel araçları kullanmak – Matematiksel sonuçların belirlenmesindeki süreçlerin ve işlemlerin yürütülmesine yardımcı olabilecek çeşitli uygun araçları bilmek ve uygun şekilde kullanmak – Matematiksel çözümün makullüğünü ve bahsi geçen problem bağlamında, bu çözümün sınırlarını ve kısıtlarını saptamak için matematiksel araçlar kullanmak

Kaynak: OECD, 2019.

Tablo 1’de sıralanan yeterlikler soru maddelerinin içeriklerinin düzenlenmesiyle, örneğin, sorunun doğrudan akıl yürütme ve kanıt gösterme (reasoning and argumentation), durumları matematiksel olarak formüleleştirme (formulating situations mathematically) gibi bilişsel becerileri ölçmesinin yanında, açık uçlu soru maddelerinin kullanılmasıyla da, öğrencinin soruya nasıl yaklaştığı, hangi problem çözme stratejilerini kullandığı ya da soruyla ilgili hangi yanlış anlamalara sahip olduğu gibi bilişsel beceriler de ölçülebilmektedir (Lie, Taylor ve Harmon, 1996). Diğer bir ifadeyle, soru içeriği düzenlemesi veya test prosedürlerinin çeşitlendirilmesiyle (ör., çoktan seçmeli sorular yanında açık uçlu sorular kullanarak), öğrencinin bu testlerdeki bilişsel ve üstbilişsel performanslarının belirlenmesi mümkün olabilmektedir. Üst düzey bilişsel becerilerin belirlenmesi konusunda, örneğin, öğrencilerin okuma becerilerinin üstbilişsel becerileriyle ilişkili olduğu bulunmuştur (Myers ve Paris, 1978; White ve Frederiksen, 2005). Ek olarak, Ardelt, Shiefele ve Schnieder (2001) da, PISA 2000 testindeki okuduğunu anlama hakkındaki üstbilişsel bilginin okuma becerisiyle (test puanı ile) yüksek düzeyde ilişkili olduğunu göstermişlerdir. Devam eden çalışmalar, örneğin, 2009 PISA testinde de aynı kanıtları sunmuştur (bkz., Ardelt ve Schneider, 2015). Alternatif bir değerlendirme yöntemi olarak, senaryo-temelli üstbilişsel bilgi testi prosedürü de (Handel, Ardelt ve Weinert, 2013), öğrencilerin öğrenme ve problem çözme stratejileri gibi üst düzey bilişsel becerilerini ölçmek için kullanılabilmiştir.

Yukarıda bahsedilen değerlendirme yöntemlerine rağmen, ülkelerarası standart olarak kullanılan PISA testi araştırmalarında genellikle öğrencilerin bilişsel özellikleri testte puanlanırken üstbilişsel becerileri ise ayrı birer ölçek üzerinde değerlendirilmiştir (bkz., Maag Merki, Ramseier ve Karlen, 2013; Wirth ve Leutner, 2008). Bu nedenle, ilgili araştırma başlıklarında, öğrencilerin üstbilişsel becerilerini ölçmek için ayrı bir ölçeğe ihtiyaç duymadan yürütülen çalışmalar yok denecek kadar az görünmektedir. Ancak, buna bir istisna olarak Higham’ın (2007) Akademik Yeterlik Sınavı’ndaki (Scholastic Aptitude Test [SAT]) üstbilişsel beceriyi incelediği çalışması gösterilebilir. Bu çalışmasında Higham, STK temelinde önerdiği hesaplama yöntemini, Koriat ve Goldsmith’in (1996) önerdiği

hesaplama yöntemleriyle karşılaştırmıştır. Çalışma sonuçları, cevaplayıcıların verdikleri yanıtlarda daha fazla doğru yanıt vermek üzere soruya yanıt verip vermemeyi seçmek olarak tanımlanabilecek “kesinlik düzenlemesi” (regulation of accuracy) becerisinin, Koriat ve Goldsmith’in önerdiği eşik teorisine göre STK ile çok daha keskin şekilde hesaplanabileceğini göstermiştir.

Bu çalışmada, Higham’ın (2007) önerdiği bu hesaplama yönteminin ilk kez PISA testi değerlendirmesinde kullanılması amaçlanmıştır. Devam eden bölümde detaylandırılan Sinyal Tespit Kuramı çerçevesindeki bu hesaplama yöntemiyle, özetle PISA testindeki matematik becerilerinin öngörülen beceriler temelinde puanlanmasının yanı sıra, öğrencilerin doğru ve yanlış yanıtlarını ne kadar iyi şekilde ayırt edebildiklerine karşılık gelen üstbilişsel izleme performanslarının saptanması ve bu performansın Bilişsel Tanı Modeli’yle elde edilecek becerilerle (örtük sınıflarla) ilişkisinin ortaya çıkarılması amaçlanmaktadır.

Bilginin zihinde mi üretildiği (hatalı bilgi) yoksa gerçekten yaşanan bir olaya ait bir bilginin depolandığı bellek alandan mı çağrıldığı (doğru bilgi) konusuna odaklanan “kaynak izleme” (source monitoring) veya yaşa bağlı bilişsel becerilerdeki değişimleri inceleyen sayısız araştırma bulunmaktadır (ör., Baltès, Staundinger ve Lindenberger, 1999). Ancak, bu araştırmaların üstbilişsel izleme performansını doğrudan incelenmek ya da belirlenmek istenen bir değişken olarak ele almadığı anlaşılmaktadır. Diğer bir ifadeyle, neden bazı kişilerin bu beceride daha iyi, bazılarının ise bu beceride daha kötü performans sergiledikleri konusu açık görünmemektedir. Bu anlamda, Dunlosky ve Tauber’in (2001) öne sürdüğü gibi, kısa süreli bellek, epizodik (zamansal) bellek, problem çözme becerilerinin azalması ve yeni bağlantıların kurulmasında (yani öğrenmede) güçlük gibi bilişsel becerilerde yaşla ilintili olarak azalış gözlemlenirken, üstbilişsel izleme performansında da yine yaşa bağlı olarak bir azalış görülmektedir. Ancak, azalan bu bilişsel becerilerin, izleme performansının azalmasıyla eş zamanlı olarak gözlemlenmesi nedeniyle, bilişsel becerilerin azalmasının üstbilişsel becerilerin azalışından kaynaklandığı sonucunu çıkarmak mümkün görünmemektedir. Bu nedenle, özellikle eğitim alanındaki diğer bilişsel yeterlikler konusunda yürütülen çalışmalardan farklı olarak bu araştırmada, izleme performansının ilişkili olabileceği diğer alt becerilerin neler olabileceği, aynı test içinde ve tek bir ölçümle, öğrencilerin bilişsel yeterlikleri BTM yöntemiyle sınıflandırılırken aynı öğrencilere ait üstbilişsel izleme performanslarının da ne düzeyde olduklarının hipotez testleri ile karşılaştırılarak incelenmesi amaçlanmıştır.

Devam eden bölümler, araştırmada kullanılan sinyal tespit kuramı ve bilişsel tanı modelini detaylandırmaktadır.

Sinyal Tespit Kuramı ve Üstbilişsel İzleme

İlk kez Flavell (1979) tarafından, kısaca “bilginin bilgisi” olarak tanımlanan üstbiliş, son yıllarda giderek artan sayıda araştırmacının ilgisini çekmiştir. Araştırmacılar, öğrenme kolaylığı (ease of learning [EOL]), öğrenme kararı (judgment of learning [JOL]) veya bilme hissi (feeling of knowing [FOK]) gibi üstbilişsel kararları bir takım standart tekniklerle ölçebilmektedir. Bütün bu çalışmaların ortak özelliği ve sonuçlarının bize sunduğu en temel çıktı ise, bireylerin kendi bilgileri hakkında ne düzeyde farkındalığa sahip olduklarını görgül olarak ölçümleyebilmektir. Bu anlamda, bilişsel becerilerin incelendiği çalışmalarda, üstbilişsel performansların da araştırılması, bireylerin yalnızca hangi bilgileri doğru bilip bilmediğini araştırmaktan öte (ör., testteki doğru yanıt sayısı), sahip oldukları doğru ya da yanlış yanıtlar veya bilgilerin doğruluğu hakkında ne düzeyde farkındalığa sahip olduklarını da belirleyebilmek için oldukça önemlidir. Diğer bir deyişle, bilgiye sahip olmaktan ya da yalnızca doğru bilmemekten öte, raporlanan bilgilerin doğru ya da yanlış olduklarına dair farkındalığın ne düzeyde olduğunu ölçmek, araştırmacılara bireylerin doğru yanıt sayılarını hesaplamadan ötesinde veriler sağlayabilmektedir.

Üstbilişin ölçülmesi, çeşitli tekniklerle yapılabilmesine rağmen (ör., öğrenme kolaylığı, öğrenme kararı, bilme hissi, vb. kararlarına ait ölçümler), günümüzde yaygın olarak Green ve Swets’in (1966) önerdiği Sinyal Tespit Kuramı ile, özellikle Tip-2 olarak adlandırılan STK ile de standart şekilde ölçülebilmektedir. STK’nın varsayımlarına göre, katılımcıların “isabet oranları (hit rate)” ve “hatalı

uyarı (false alarm)” hesaplanarak, bu katılımcılara ait üstbilişsel izleme performansları ve yanıtlarını verirken ne düzeyde katı ya da esnek davrandıklarını işaret eden “raporlama kriteri (report criterion)” ölçülebilmektedir. Fiziksel uyarıcı ve bunun duyum ve algı ile olan ilişkisine odaklanan “psikofizik” (Luce ve Krumhansl, 1988) bakış açısına dayanan Sinyal Tespit Kuramı (Green ve Swets, 1966), Tip-1 ve Tip-2 olarak iki türde değerlendirilebilir (ör., Higham, 2002; Higham ve Tam, 2005). Standart olarak ölçülen Tip-1 STK’ya göre, katılımcıdan, örneğin, beyaz ses (white noise) olarak tanımlanan gürültülü bir ses içinde gizlenen sinyali tespit edip edemediğini “evet”/“hayır” olarak belirtmesi istenir. Sinyal, bazı denemelerde gürültünün içinde gerçekten de bulunurken, bazı denemelerde böyle bir sinyal bulunmaz. Kurama göre, sinyalin bulunduğu ve bulunmadığı denemelerin iki normal dağılım oluşturdukları varsayılır ve katılımcının sahip olduğu raporlama kriterinin üzerinde kalan denemeler için sinyal varlığına ‘evet’ yanıtı verirken, bu eşğin altında kalan denemeler için sinyal varlığına ise ‘hayır’ yanıtı verdiği varsayılır. Sinyalin gerçekte bulunduğu ve bulunmadığı durumlar (sinyal: var-yok) ile katılımcının bu iki durum için verdiği “evet”/“hayır” yanıtlarına ait olasılıkların türü belirlenebilir; bkz. Tablo 2. Buna göre, *ısabet oranı*, katılımcının “Evet (var)” yanıtı verdiği denemelerdeki sayının, gerçekten bir sinyalin bulunduğu bu denemelerin toplam sayısına oranıdır. *Hatalı uyarı oranı* ise, katılımcının yine “Evet” yanıtı verdiği denemedeki sayının, gerçekte böyle bir sinyalin bulunmadığı denemelerin toplam sayısına oranıdır. Ancak katılımcılar, bir sinyal bulunmazken uygun şekilde beklendiği üzere “Hayır (yok)” yanıtını da verebilirler. Bu denemelerde ölçülen oran ise, *doğru ret* (correct rejection) olarak tanımlanırken, katılımcılar yine gerçekte bir sinyal bulunmasına rağmen yanıtlarını “yok” olarak verebilirler. Son duruma ait olan oran, bu türdeki denemeler için *kaçırma* (miss) olarak tanımlanır (Abdi, 2007). Özetle, Tip-1 olarak adlandırılan bu ölçümde, katılımcının aslında bir makine benzerinde pasif bir rol edinerek sinyalleri doğru şekilde tespit edip etmediğine odaklanılır ve bu sinyal tespit yöntemi, uyarıcı-bağımlı (stimulus-contingent) olarak nitelendirilir (Higham ve Tam, 2005).

Tablo 2. Tip-1 Sinyal tespit kuramına göre elde edilebilecek dört olası yanıt

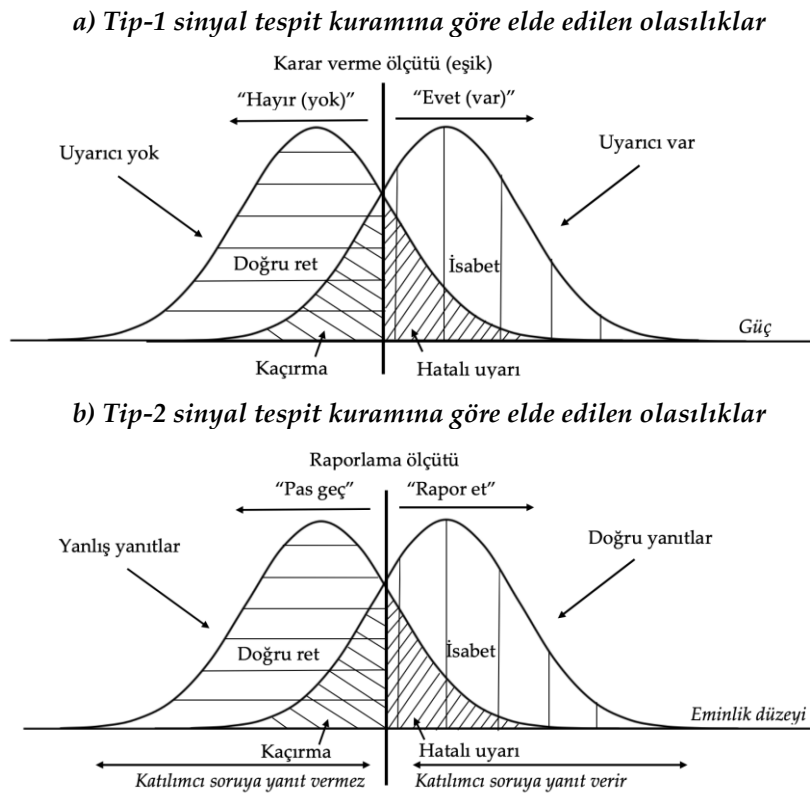
	Karar (katılımcının yanıtı)	
Gerçek	“Evet (var)”	“Hayır (yok)”
Sinyal var	İsabet (hit) (a)	Kaçırma (miss) (b)
Sinyal yok	Hatalı uyarı (false alarm) (c)	Doğru ret (correct rejection) (d)

Not. İsbet oranı= $a/(a+b)$; hatalı uyarı oranı= $c/(c+d)$; kaçırma oranı= $b/(a+b)$; doğru ret oranı= $d/(c+d)$

Kaynak: Abdi, 2007.

Tip-2 sinyal tespit kuramına göre ise, yukarıda bahsedilen oranların isimleri aynı kalmakla birlikte, ölçülen değerlerin özellikleri farklıdır. Buna göre, katılımcı örneğin yöneltilen bir soru için yanıtlarını oluştururken zihninde yanlış ve doğru yanıt olabilecek listeler üretir. Diğer bir deyişle, varsayımsal olarak bir bilginin doğrudan ve tıpkı bir vakum gibi depolama alanından çağrılmasından ziyade, yanıt vermeden önce yanlış ve doğru yanıtları içeren yanıtlarının bilince aktığı varsayılır (bkz., Bahrick, 1970; Kintsch, 1970; Watkins ve Gardiner, 1979). Tip-2 STK, yanlış ve doğru yanıtların sinyalleme güçleri ve dolayısıyla eminlik düzeyleri bakımından normal dağıldıklarını varsayarak, bu dağılımların ortalama değerleri arasındaki farkın hesaplanmasıyla (ör., d'), bireyin doğru ve yanlış yanıtları ne düzeyde ayırt edebildiklerine dair farkındalıklarını ölçmeyi hedeflemektedir. Tip-2 STK’daki raporlama kriteri ise, katılımcının ürettiği yanıtları, bu kriterin üzerindeyse rapor etmek istemesini belirler. Diğer bir deyişle, üretilen yanıtlar eğer bu kriterin üzerindeyse, katılımcı yanıtının doğru olduğunu düşünmekte ve rapor etme (soruya yanıt verme) eğilimindedir. Bu varsayımdan da çıkarsanabileceği gibi, standart olarak katılımcıya doğru yanıtı vermeden önce, örneğin soruyu “yanıtlayacağım” ya da “yanıtlamayacağım” (çalışmalarda sıklıkla kullanıldığı gibi “raporlayacağım” ya da “pas geçeceğim” [report/pass]) şeklinde bir yanıt vermesi istenir. Katılımcı “yanıtlamayacağım” seçeneğini vermiş olsa da, katılımcıdan sorunun yanıtına ait en iyi tahmininin ne olduğunu belirtmesi istenir. Bu durumda, örneğin katılımcı bazen yüksek bir raporlama kriteri koymuş ve gerçekte doğru olabilecek yanıtı için “yanıtlamayacağım” seçeneğini belirtmiş olabilir. Yani, katılımcının yanıtından yalnızca yüksek derecede emin olduğu sorulara yanıt verme eğiliminde olduğu düşünülebilir. Son olarak,

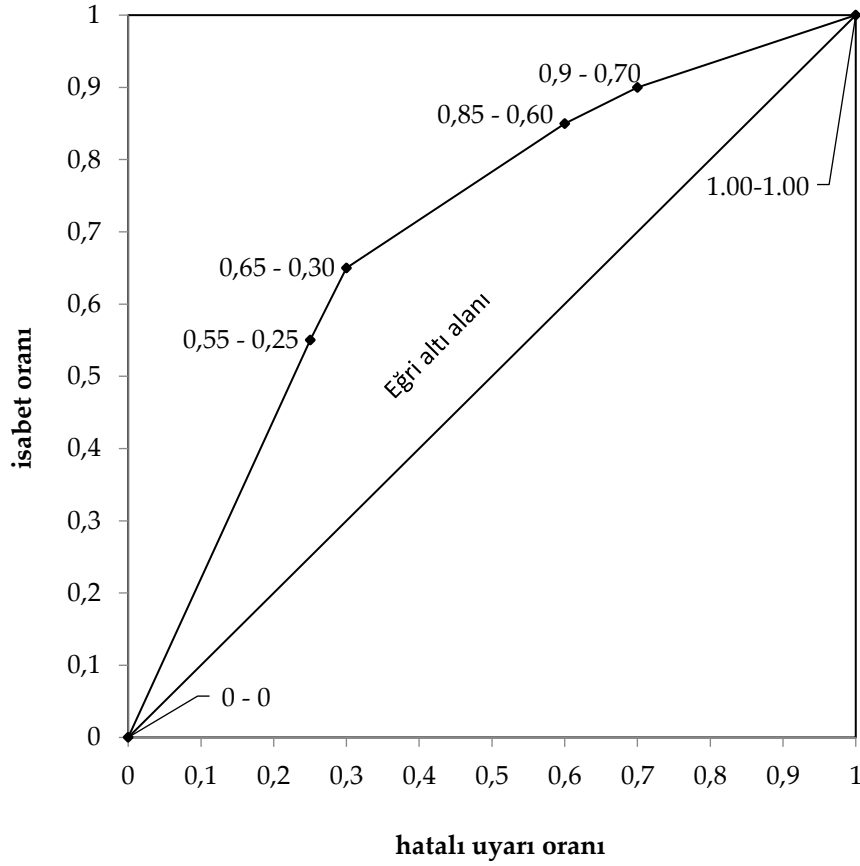
katılımcıdan verdiği yanıtın doğru olduğundan ne kadar emin olduğunu Likert-tipi bir ölçekte belirlemesi istenir. Özetle, Tip-2 STK için oluşturulacak çapraz çizelge (contingency table) 2(yanıtlama: raporlama - pas geçme) x 2 (üretilen aday yanıt: doğru - yanlış) şeklindedir (bkz., Higham ve Tam, 2005, s. 599). Yanlış üretilen yanıtlara ait dağılım tarafında (Şekil 1(b) soldaki dağılım) düşük eminlik derecesinden başlamak üzere, doğru üretilen yanıtlara ait dağılım tarafındaki (Şekil 1(b) sağdaki dağılım) eminlik düzeyleri yüksek değerlere sahiptir. Özetle, bireyin doğru ve yanlış yanıtlarını ayırt edebilme becerisi, doğru ve yanlış olarak üretilen yanıtlara ait dağılımların birbirinden ne kadar uzak şekilde dağıldığına ve raporlama kriterinin katılımcı tarafından hangi düzeye yerleştirildiğine bağlıdır (katı veya esnek raporlama kriteri [stringent vs. lenient response criterions]). Tüm bu varsayımlar temelinde özetlemek gerekirse, Tip-1 STK uyarıcı-bağımlı iken, Tip-2 STK ise tepki-bağımlıdır ([response-contingent]; Higham, 2002, 2011).



Şekil 1. Tip-1 ve Tip-2 Sinyal Tespit Kuramı'na göre elde edilebilecek olasılıklar.

(Kaynak: Higham ve Arnold, 2007)

Üstbilişsel izleme becerisinin düzeyini belirlemek, zihinde üretilen doğru ve yanlış yanıtlara ait iki dağılım arasındaki uzaklığın isabet ve hatalı uyarı oranları belirlendikten sonra hesaplanması mümkündür. Araştırmacılar, yanlış ve doğru olarak üretilen yanıtlara ait iki dağılımın tepe noktaları arasındaki uzaklığı, örneğin " d' " (d' -prime) adıyla hesaplayabilmektedir ($d' = Z(\text{isabet oranı}) - Z(\text{hatalı uyarı})$). Bu değer, grafiksel olarak gösterildiğinde, *karar vericinin etkinlik özellikleri* (receiver operating characteristics [ROC]) yöntemiyle de gösterilebilir. Bu amaçla, katılımcının her bir yığılmalı eminlik değerindeki ve "0,00" ve "1,00" arasında değişen isabet ve hatalı uyarım oranları hesaplanarak, bunların kesişim değerleri saçılım grafiğine işaretlenir; bkz., Şekil 2. İşaretlenen noktaların birleştirilmesi ile, hiçbir bilinçli ayrımı işaret etmeyen (diğer bir ifadeyle %50 olasılıkla zaten doğru cevaplanabilecek raporlamayı ifade eden) varsayımsal çizginin üzerinde kalan alan hesaplanabilir. Bu alan, ne kadar büyük ve pozitif bir değere sahipse, katılımcının alan altı eğrisi değeri (dolayısıyla, d' değeri de) -yani doğru yanıtlarını doğru, yanlış yanıtlarını da gerçekten yanlış olarak fark ettiğine dair puanı- o kadar yüksektir. Eksi bir değer ise, katılımcının doğru yanıtlarını yanlış, yanlış yanıtlarını ise doğru olarak düşündüğünü ifade eder. Şekil 2 farazi bir katılımcının eğri altı alanının hesaplanmasında kullanılabilecek bir ROC eğrisini göstermektedir.



Şekil 2. Yanıtların doğrulukları hakkındaki eminlik düzeyi “1” ve “5” arasında değişen bir testte (1=“hiç emin değilim”; 5=“kesinlikle eminim”), varsayımsal bir katılımcıya ait eğri altı alanının hesaplanmasında çizilen “karar vericinin etkinlik özellikleri” (receiver operating characteristics [ROC]) grafiği.

Şekil 2’ye göre, hesaplanan eğri altı alan, kırık çizgilerle birleştirilen eğrinin köşegenel çizgiyle arasında kalan alandır. Birikimli eminlik değerlerindeki oranlara ait kesişimler, en yukarıdaki kesişim noktasında “1 ve üzeri” eminlik değerlerine ait olup (ör., 1,00 isabet oranı ve 1,00 hatalı uyarı oranı), diyagramın yukarıdan aşağı sayıldığındaki ortaya çıkan 5. kesişim noktası ise “yalnızca 5” eminlik değerlerinde verilen yanıtlar göz önünde bulundurularak hesaplanan oranların kesişimidir (ör., 0,55 isabet oranı ve 0,25 hatalı uyarım oranı). Diyagramdan da çıkarsanabileceği üzere, eminlik değerlerinde kullanılacak Likert-tipi ölçekte ne kadar fazla değer bulunursa, alan hesaplamasında ortaya çıkan eğri o denli daha az kırık şekilde olacak ve izleme performansı da daha kesin olarak ölçümlenebilecektir.

Bilişsel Tanı Modeli

Araştırmada kullanılan BTM, test maddeleri ve testin ölçtüğü özellikler konusundaki önsel (a priori) bilgileri kullanarak, öğrencilerin ölçülen özelliklere sahip olup olmama durumlarını en az hata ile sınıflama düzeyinde belirlemeyi amaçlamaktadır. Örtük sınıf analizleri (latent class analyses) ve çok boyutlu madde tepki kuramı (Item-Response Theory [IRT]) yaklaşımlarını temel alan BTM’de, testin ölçtüğü özelliklerin test maddeleri tarafından ne düzeyde temsil edildiği test geliştirme sürecinde belirlenir. Analizler sonunda, öğrenciler sahip oldukları özelliklere göre yetenek sınıflarına ayrılırlar. Özetle, bu analizlerdeki temel amaç, öğrencilerin hangi örtük sınıfa dahil olduklarını en az hata ile belirlemektir (Leighton ve Gierl, 2007; von Davier, 2014).

Haertel (1989) tarafından geliştirilen DINA model (Deterministic Input Noisy and Gate) (Haertel, 1989), çoğu Bilişsel Tanı Modeli gibi ikili yetenek modellerine benzeyen bir örtük sınıf analizidir (Junker, 1999; Junker ve Sijtsma, 2001; MacReady ve Dayton, 1977). Model, madde özellik ilişkisini temel alır ve modelin iyi işleyebilmesi, testteki her maddenin doğru cevaplanması için gerekli

olan özelliklerin doğru belirlenmesine bağlıdır (de la Torre ve Chiu, 2015; de la Torre ve Lee, 2010). Özetle, geliştirilen ya da uygulanan bir test için belirlenen örneğin "k" sayıdaki özelliği temel olarak "2^k" sayıda örtük sınıf belirlemektedir. Örneğin, sadece üç özelliğin ölçüldüğü bir testte, testi alan bireyler toplam 8 örtük sınıf içinde sınıflanabilir. Üç özellik için olası sınıflar, "000", "100", "010", "001", "110", "011", "101" ve "111" şeklinde sıralanabilir. Hiçbir özelliğe sahip olmayanlar ilk sınıfa dahil olurken ("000"), sadece birinci ve üçüncü özelliğe sahip olan bireyler 7. sınıfa ("101") yerleşirler. Bu analizde önemli olan nokta, bir cevaplayıcının bir özelliğe sahip olup olmadığını belirleme aşamasının olasılık temeline dayanmasıdır. Öğrencinin özellik bakımından "0" sınıfında mı yoksa "1" sınıfında mı olacağı bir olasılık değeridir ve bu değer, araştırmacı tarafından değiştirilebileceği gibi, genellikle kullanımda ".50" eşiği temel alınır. Diğer bir ifadeyle, öğrencinin belirlenen özelliğe sahip olma olasılığı ".50" değerinin altında kalırsa "0" sınıfına (bu özelliğe sahip değil), üstünde ya da aynı değeri alırsa "1" sınıfına dahil olur (bu özelliğe sahip) (Başoğlu, 2012, 2014; de la Torre, 2008; de la Torre ve Douglas, 2004).

Araştırmanın bağımsız değişkenlerinden biri olan matematiksel yeterlikler sürekli değil kategorik bir değişken olarak analizlere alınmış ve öğrencilerin bu yeterliklere sahip olup olmaması ile ilgili temel kararlar BTM yöntemlerinden biri olan DINA model kullanılarak verilmiştir.

Yöntem

Katılımcılar

Araştırma örneklemini, İzmir İli Bornova ilçesinde yer alan bir ortaokulun bütün 6. sınıf öğrencilerini kapsamaktadır. Okuldaki mevcut 7 şubeden toplam 230 (110 erkek, 120 kız) öğrenci araştırmaya katılmıştır. Test analizleri ve DINA model için madde parametreleri ve öğrenci yetenek sınıfları belirleme çalışmaları, 230 öğrenci üzerinden gerçekleştirilmiştir. Üstbilis analizleri ise, örneklemden tamamen tesadüfi olarak belirlenen (58 erkek, 72 kız) 130 öğrenci üzerinden gerçekleştirilmiştir. Üstbilis ölçümleri alınan öğrencilerden 10 tanesinin AUC hesaplanması için gerekli olan tüm soruları eksiksiz cevapla koşulunu sağlayamadığı için bu katılımcılar karşılaştırmalı analize dahil edilmemiştir. Böylece AUC analizleri geri kalan 120 öğrenci için yapılmıştır. Üstbilis ölçümü yapılan grubun şubelere dağılımı homojendir. Test örneklemini (ortalama=2,41; ss=2,28) ile deney örnekleminin (ortalama=2,44; ss=1,83) test performansları arasında anlamlı bir fark bulunmamaktadır.

İşlem

Araştırma TÜBİTAK 115K531 No.lu projesinin uygulamaları kapsamında, genel örneklemin pilot grubu üzerinden gerçekleştirilmiştir. Süreç bütünüyle İzmir İl Milli Eğitim Müdürlüğü ile koordineli olarak yürütülmüştür. Okul yönetiminin kontrolünde, öğrencilere veli onay formları dağıtılmış ve çalışmaya velileri onay veren öğrenciler dahil edilmiştir. Çalışmaya katılmayan az sayıdaki öğrenciler, ölçüm sırasında sınıfta bulunmuşlar ancak ölçüm sırasında onlardan dersle ilgili okumalar yapmalarını istenmiştir. Aşağıdaki "Ölçme Aracı" bölümünde detayları verilen ve 12 maddeden oluşan PISA eşdeğerli matematik testi, toplam 230 6. sınıf öğrencisine, kendi sınıflarında ve eş zamanlı olarak uygulanmıştır. Proje asistanları, test sırasında test yürütücüsü ve gözetmenlik görevi alan sınıf öğretmenlerine soru kitapçıklarını ve yönergeleri Milli Eğitim Bakanlığı'na bağlı memurlar aracılığıyla ulaştırmıştır. Toplanan veriler, ilgili memurlar tarafından her bir test kitapçığının kapalı zarflarla konulması sonrası posta yoluyla proje ekibine ulaştırılmıştır.

Ölçme Aracı

Araştırmada, 12 maddeden oluşan ve PISA matematik yeterliklerini ölçmeyi amaçlayan 6. sınıf düzeyinde matematik testi kullanılmıştır. Testin geliştirme ve ölçülen becerilerin belirlenmesi süreci Matematik Eğitimi alan uzmanları ile birlikte yürütülmüştür. Araştırma kapsamında uyarlanan PISA yeterlik alanları proje araştırmacıları, danışmanlar ve alan uzmanı öğretmenlerle birlikte madde türleri, konu ve sınıf düzeyi göz önünde bulundurularak dört başlık altında toplanmıştır. "İletişim ve ilişkilendirme", "matematikleştirme", "akıl yürütme ve strateji geliştirme", "sembolik ve teknik dil

kullanımı” olarak belirlenen bu dört yeterliğe ilişkin tanımlamalar şu şekilde açıklanmıştır (Başokçu, 2019):

İletişim ve İlişkilendirme: İletişim becerisi; matematiksel dili, günlük dil ve sembollerle ilişkilendirme, matematiksel düşüncelerin doğruluğunu ve anlamını yorumlama gibi özellikleri kapsar. İlişkilendirme becerisi, matematik kavramlarının kendi aralarında, diğer disiplinlerle ve günlük hayatla ilişkilendirilmesini kapsamaktadır.

Matematikleştirme: Bir problemi kesin matematiksel formda ifade etmeye yönelik modelleme, yapısal gösterim, varsayımlarla tanımlama, formülleştirme, kurulan bir yapı ya da modele yönelik matematiksel çıktılar elde edilmesi ve yorumlanması gibi belirli eylemleri ifade etmektedir.

Akıllı Yürütme ve Strateji Geliştirme: Akıllı yürütme, eldeki bilgilerden hareketle matematiğin kendine özgü araç (semboller, tanımlar, ilişkiler, vb.) ve düşünme tekniklerini (tümevarım, tümdengelim, karşılaştırma, genelleme, vb.) kullanarak yeni bilgiler elde etme sürecidir. Strateji geliştirme, problem çözmede matematiksel bilgi ve becerilerin kullanılmasına yönelik bir plan ya da stratejinin seçilmesi ya da tasarlanmasını ifade etmektedir.

Sembolik ve Teknik Dil Kullanımı: Matematik okuryazarlığı açısından sembolik ve teknik dilin kullanılması yeteneği, matematiksel kurallarla ifade edilen bir matematiksel içeriğe yönelik olarak sembol gösterimlerini anlama ve yorumlama davranışlarının bir bütünüdür.

Test maddelerine ait parametreler, belirlenen test maddelerinin ölçtüğü ve yukarıda detayları verilen matematiksel beceriler ile ölçme aracına ait DINA model madde parametreleri Tablo 3’de gösterilmiştir (sırasıyla, Tablo 3(a), 3(b), 3(c)).

Tablo 3. Ölçme aracına ait madde parametreleri (a), DINA modelde belirlenen test maddelerinin ölçtüğü matematiksel beceriler (b; 0=madde beceriyi ölçmez; 1=madde beceriyi ölçer) ve ölçme aracına ait DINA model madde parametreleri (tahmin ve kaydırma) ve parametrelere ait standart hata puanları (c).

a) Ölçme aracına ait madde parametreleri												
	Test maddeleri											
	1	2	3	4	5	6	7	8	9	10	11	12
Güçlük	0,28	0,12	0,02	0,37	0,33	0,11	0,24	0,07	0,35	0,28	0,05	0,18
Ayır edicilik	0,46	0,26	0,06	0,58	0,47	0,27	0,37	0,16	0,58	0,42	0,12	0,35
Madde-test korelasyonu	0,47	0,51	0,51	0,53	0,45	0,68	0,48	0,55	0,55	0,48	0,65	0,5
b) Test maddelerinin ölçtüğü matematiksel beceriler												
İletişim ve ilişkilendirme	1	1	0	1	1	0	0	0	0	1	1	1
Matematikleştirme	0	0	1	0	0	1	1	1	1	0	0	0
Akıllı yürütme ve strateji geliştirme	1	1	1	0	0	0	1	1	1	0	1	0
Sembolik ve teknik dil kullanımı	0	0	1	1	1	1	0	0	1	0	1	0
c) Ölçme aracına ait DINA Model madde parametreleri ve standart hataları												
Tahmin (G-par)	0,21	0,04	0,02	0,16	0,33	0,07	0,12	0,04	0,31	0,10	0,03	0,17
Tahmin standart hata	0,04	0,03	0,01	0,05	0,04	0,03	0,05	0,02	0,04	0,06	0,01	0,05
Kaydırma (S-par)	0,21	0,25	0,54	0,06	0,26	0,46	0,41	0,45	0,38	0,24	0,49	0,32
Kaydırma standart hata	0,10	0,10	0,05	0,08	0,07	0,11	0,14	0,06	0,13	0,07	0,07	0,05

Test maddelerin güçlük ortalaması 0,20, ayırt-edicilik ortalaması 0,34 ve madde test korelasyonları ortalaması 0,53 olarak hesaplanmıştır; bkz. Tablo 3(a). Test için hesaplanan KR-20 güvenirlik katsayısı 0,73 olarak hesaplanmıştır. Bu değer, 12 maddelik bir test için yüksek düzeyde güvenirliğe işaret etmektedir. Aynı zamanda testin güvenirliği ve geçerliğini Madde Tepki Kuramı (MTK) analizleri yardımı ile de incelenmiştir. Analizler İki-Parametrelili Lojistik Model kullanılarak gerçekleştirilmiştir. Oğiv fonksiyon sergileyen test bilgi fonksiyonu ve madde karakteristik eğrileri incelenmiş olup buna göre, oluşturulan testin ölçmeyi amaçladığı değişkeni yüksek düzeyde ayırt edici ve güvenilir olarak ölçebildiği gözlenmiştir. Bunlara ek olarak, araştırmada kullanılan DINA modelin çok boyutlu yapısı göz önüne alındığında test ayırtediciliği ve geçerliği ile ilgili sunulan Klasik Test Kuramı (KTK) ve MTK kanıtların bir karşılaştırma kriteri olarak kullanılması gerekmektedir.

Maddelerin ölçülen beceriler ile ilişkilendirilme süreci uzman görüşleri göz önüne alınarak belirlenmiş ve Tablo 3(b)'de verilen Q matris hazırlanmıştır. Bu süreç ölçek geliştirme sürecindeki maddeler ile boyutların kuramsal olarak eşleştirilmesi yaklaşımına benzemektedir. BTM modellerinin farklılığı ise Q matriste bir maddenin birden fazla özellik (boyut) ile eşleştirilebilmesidir. Oluşturulan bu matris BTM analizleri için kullanılan temel önsel ve kuramsal ilişkileri göstermektedir. Bu matris kullanılarak DINA model madde parametrelerini ve testi alan bireylerin örtük sınıflarını belirlemektedir. Tablo 3(b)'de gösterildiği gibi, sonuçlar toplam 7 maddenin "iletişim ve ilişkilendirme", 5 maddenin "matematikleştirme", 7 maddenin "akıl yürütme ve strateji geliştirme" ve 6 maddenin ise "sembolik ve teknik dil kullanımı" becerisini ölçtüğünü göstermektedir. BTM'nin ayırt edici özelliklerden biri olarak, bir maddenin birden fazla özellik ile ilişkilendirilebilmesi durumu Tablo 3(b)'de gösterilmektedir. Diğer bir ifadeyle, modelin çok boyutlu yapısının bir göstergesi olarak, testte sadece 2 madde yalnızca "bir" özelliği ölçerken, 7 madde "iki" özelliği, 3 madde ise "üç" özelliği ölçmektedir. Tablo 3(c) incelendiğinde ise, "tahmin" (G-par) parametrelerinin düşük olduğu görülmektedir. "Kaydırma" (S-par) parametrelerinin ise özellikle güçlük düzeyi yüksek maddelerde yükseldiği gözlenmektedir. Bu durum, BTM için beklenen bir sonuç olmakla birlikte, genel parametre değerleri ve bunların ortalamaları göz önüne alındığında modelin iyi uyum gösterdiği görülmektedir. Aynı zamanda model uyumunun bağıl indeksleri olan AIC (Akaike Information Criterion [Akaike Bilgi Kriteri]) ve BIC (Bayesian Information Criterion [Bayes Bilgi Kriteri]) sırasıyla 2535,38 ve 2669,46 olarak hesaplanmıştır.

Katılımcıların üstbilişsel izleme performansları ise, Şekil 2'de gösterilen "eğri altı alan" (area under the curve [AUC]) yöntemiyle her bir katılımcı için ayrı ayrı olarak hesaplanmıştır. Katılımcılardan, her bir test maddesi için soruyu okuyup, bu soruya önce "soruyu çözebileceğimi düşünüyorum" ya da "soruyu çözebileceğimi düşünmüyorum" seçeneğinden birini belirtmesi istenmiştir. Katılımcılardan, soruyu çözemeyeceğini düşünseler de yine de en iyi tahminleri yapmaları istenmiştir. Soruyu çözdükten sonra (ya da soruyu pas geçseler de doğru olduğunu düşündükleri yanıt seçtikten sonra) her bir yanıt için bu yanıtların doğru olduklarından ne kadar emin olduklarını 5-li Likert-tipi ölçek üzerinde belirtmeleri istenmiştir (1= "hiç emin değilim"; 5 = "kesinlikle eminim"). Bütün yığılmalı eminlik derecesindeki ("1" ve üzeri, "2" ve üzeri, "3" ve üzeri, "4" ve üzeri eminlikler ve yalnızca "5" eminlik düzeyindeki) isabet oranlarının ve hatalı uyarım oranlarının kesişimleri her katılımcı için hesaplanarak, katılımcıların karar vericinin etkinlik özellikleri (ROC) eğrileri oluşturulmuş ve bu eğrinin Şekil 2'de gösterilen diyagonal çizgiye kadar altında kalan alan hesaplanmıştır. Örneğin, "1 ve üzeri" eminlikteki isabet oranı, katılımcının yanıt vermeyi seçtiği ve doğru olarak yanıtladığı ve bunları herhangi bir eminlik düzeyi ile derecelendirdiği yanıtlarının toplam sayısının, yanıt vermek istediği ve pas geçtiği doğru yanıtlarının toplam sayısına oranıdır. "2 ve üzeri" eminlik düzeyindeki isabet oranı hesaplanırken de katılımcının eminlik düzeyi için "2" ya da daha yüksek bir derecelendirme yaptığı doğru yanıtları hesaba katılmıştır. Hatalı uyarım oranı ise, örneğin yine "1 ve üzeri" eminlik düzeyi için, katılımcının yanıt vermeyi seçtiği ancak yanlış cevap verdiği ve herhangi bir eminlik derecesi verdiği yanlış yanıt sayısının, yanıt verdiği ya da pas geçtiği sorulardaki

yanlış yanıtların toplam sayısına oranıdır. Aynı şekilde, örneğin “2 ve üzeri” eminlik düzeyi için bu oran hesaplanırken, yalnızca “2” eminlik düzeyi ya da daha daha yüksek eminlik düzeyi ile derecelendirdiği yanıtlar hesaba katılmıştır (ör., *d'*, yanıt verme yanlığı [reponse bias] gibi hesaplamalarda kullanılan formüller için, bkz., Higham ve Tam, 2005, s. 599). Her bir katılımcı için, tüm yığılmalı eminlik düzeylerindeki isabet oranları ve hatalı uyarım oranları yukarıda bahsedilen formüllere göre hesaplanmış ve oluşan eğri altındaki alanın (AUC) büyüklüğü, yine her katılımcı için üstbilişsel izleme performansı (diğer bir ifadeyle doğru ve yanlış yanıtlarını ayırt edebilme becerisi) olarak belirlenmiştir.

Bulgular

Altıncı sınıf öğrencilerinin matematik becerilerinin PISA matematik testiyle ölçüldüğü bu çalışmadan elde edilen veriler, daha önceden bu testin değerlendirilmesinde kullanılmamış olan örtük sınıf analizi yöntemine dayanan DINA model kullanılarak analiz edilmiştir. Analizlerde Ox Edit ve R programları için geliştirilmiş BTM paketleri kullanılmıştır (George, Robitzsch, Kiefer, Groß ve Ünlü, 2016). Üstbilişsel izleme becerisi ise, daha önceden yalnızca SAT testinde uygulanan (Higham, 2007) ve Tip-2 sinyal tespit kuramı olarak tanımlanan hesaplama yöntemiyle (alan altı eğrisi ile) elde edilmiştir (bkz. Sinyal Tespit Kuramı ve Üstbilişsel İzleme bölümü ve Şekil 1). Katılımcıların örtük sınıflara ait sonsal olasılıkları, analiz sonrası elde edilen sınıfların gözlenen olasılıkları, bu sınıflara ait testten alınan ortalama doğru sayısı ve örtük sınıfların üstbilişsel izleme performansları Tablo 4’te verilmiştir.

Tablo 4. DINA Örtük Sınıflara ait Sonsal Olasılıklar, belirlenen örtük sınıfların testten aldığı ortalama doğru yanıt sayıları ve sınıfların alan altı eğri (AUC) hesaplamalarıyla elde edilen üstbilişsel izleme performansları

Örtük sınıf *	Sonsal olasılık	Gözlenen olasılık	Ortalama doğru sayısı	Üstbilişsel izleme becerisi (AUC)
"0000"	0,0658	0.187	1,12	0,16
"1000"	0,0631	0.106	2,50	0,14
"0100"	0,0658			
"0010"	0,0658			
"0001"	0,0658	0.132		
"1100"	0,0631	0.081	1,75	0,13
"1010"	0,0538	0.074	2,50	0,18
"1001"	0,0644	0.076	2,21	0,13
"0110"	0,0572	0.052	2,21	0,22
"0101"	0,0623			
"0011"	0,0658			
"1110"	0,0476	0.084	4,67	0,34
"1101"	0,0527	0.087	3,89	0,34
"1011"	0,065			
"0111"	0,0482			
"1111"	0,0936	0.121	6,36	0,41

* Örtük sınıflar, “XXXX” şeklinde dört haneli olarak gösterilmiş ve sırasıyla “iletişim ve ilişkilendirme”, “matematikleştirme”, “akıl yürütme ve strateji geliştirme” ve “sembolik ve teknik dil kullanımı” özelliklerini göstermektedir. Kodlamalardaki “1” rakamı, sırada belirtilen özelliğe sahip olmayı gösterirken “0” ise bu becerinin olmadığını işaret etmektedir. Örneğin, “1010” kodu, “iletişim ve ilişkilendirme” ile “akıl yürütme ve strateji geliştirme” becerilerine sahipken “matematikleştirme” ve “sembolik ve teknik dil kullanımı” becerisine sahip olmayan örtük sınıfa ait koddur.

Tablo 4’te sonsal olasılık değerleri model parametrelerine göre örtük sınıfların tamamına ait değerlerdir. Ancak her örneklemede olası örtük sınıfların hepsi gözlenemeyebilir. Tablo 4’de görüldüğü gibi örneklemede 16 örtük sınıftan sadece 9 tanesi gözlenmiştir. Tablo 4’teki olasılık değerlerine göre,

örtük sınıflara ilişkin sonsal olasılıklar birbirine yakın değerler almaktadır. Bu durum, belirlenen özelliklerin birbirinden bağımsız olduğunun bir göstergesi olarak kabul edilmektedir (Chen, de la Torre ve Zhang, 2013; Huo ve de la Torre, 2014). Tabloda gösterildiği gibi, örtük sınıflarda sahip olunan özellik sayısı arttıkça (diğer bir ifadeyle, beceriye sahip olduğunu işaret eden “1” rakamının sayısı arttıkça), o örtük sınıftaki test ortalaması ve üstbilişsel izleme puanları da artış göstermektedir. Örtük sınıf analizine göre, örneğin hiçbir belirlenen beceriye sahip olmayan örtük sınıf (“0000”) testte 1,12 ortalama doğru yaparken bu becerilerin tümüne sahip olan “1111” örtük sınıfı, testte ortalama olarak 6.36 soruyu doğru yanıtlamıştır. Becerilerin hiçbirine sahip olmayan grubun üstbilişsel izleme performansı da (.16), becerilerin hepsine sahip olan gruba göre (.41) anlamlı olarak düşüktür ($t(50)=3,59, p<0,001$). Hangi belli özelliklerin, üstbilişsel izleme performansında anlamlı olarak farklılaştığını belirlemek için, özelliklere sahip olup olmama durumu ile üstbilişsel izleme becerisi arasındaki farklar, tüm beceriler ayrı ayrı olmak üzere, ilişkisiz örneklem t-test ile analiz edilmiştir; bkz. Tablo 5.

Tablo 5. Becerilere sahip olan (1) ve olmayan grupların (0) üstbilişsel izleme performanslarının (alan altı eğrisi) karşılaştırılmasına ait ilişkisiz örneklem t-testi sonuçları

	Beceriye sahip olup olmama	n	ort.	Ss	t	df	p
İletişim ve ilişkilendirme	1	60	,23	,29	0,953	118	,343
	0	60	,18	,28			
Matematikleştirme	1	40	,27	,27	1,823	118	,071
	0	80	,17	,29			
Akıl Yürütme ve Strateji Geliştirme	1	42	,30	,30	2,914	118	,004
	0	78	,15	,26			
Sembolik ve Teknik Dil Kullanımı	1	44	,23	,31	0,945	118	,347
	0	76	,18	,27			

Not. n=katılımcı sayısı, ort.= izleme performansı ortalaması, ss=standart sapma, df=serbestlik derecesi, p=alfa değeri

Tablo 5’te gösterilen t-testi sonuçlarına göre, sadece “akıl yürütme ve strateji geliştirme” özelliğine sahip olmanın üstbilişsel izleme üzerinde etkili olduğu bulunmuştur. Diğer bir ifadeyle, akıl yürütme ve strateji geliştirme özelliğine sahip olan öğrencilerin üstbilişsel izleme puanları ortalaması (.30), bu özelliğe sahip olmayan öğrencilerin ortalamasından (.15) istatistiksel olarak orta düzeyde bir etki büyüklüğünde (Cohen $d=0,55$) anlamlı düzeyde yüksektir. İletişim ve ilişkilendirme, matematikleştirme ve sembolik ve teknik dil kullanımı özelliklerine sahip olup olmama durumunda, öğrencilerin üstbilişsel izleme puanları arasında anlamlı bir fark bulunmamıştır.

Tartışma, Sonuç ve Öneriler

BTM analizleri bir testi oluşturan maddeleri testin ölçtüğü özellikler ile ilişkilendirerek bu maddelere verilen cevaplar üzerinden öğrencileri örtük sınıflara atamaktadır (Henson ve Douglas, 2005). Aynı zamanda, test ve özelliklerin yapılandırılması, belirlenen örtük sınıflar için bir ölçüm modeli de ortaya koyabilmektedir (DiBello, Roussos ve Stout, 2006). Böylece, BTM parametreleri, kurulan modelin uyumu hakkında da araştırmacılara bilgi verebilmektedir (Hu, Miller, Huggins-Manley ve Chen, 2016). Araştırma bulgularına göre, BTM parametreleri ve örtük sınıf sonsal dağılımları, matematiksel yeterlikler göz önüne alınarak kurulan ölçüm modelinin uyum düzeyinin yeterli olduğuna işaret etmektedir. Bu durum, test maddelerinin (Tablo 3(a)) ve Q matrisinin yeterli geçerliğe (Tablo 3(b)) sahip olduğu konusunda da bilgi vermektedir. Geliştirilen ve uygulanan testin psikometrik özellikleri hem klasik test kuramı hem de madde tepki kuramı yöntemleri kullanılarak incelenmiş ve test güvenilirliğine ve madde geçerliğine ilişkin kanıtlara her iki yönetime ait analizlerle de ulaşılmıştır. Bu bulgular göz önüne alındığında, araştırmada DINA model kullanılarak belirlenen sınıflama sonuçlarının yüksek düzeyde geçerli ölçümler olduğu görülmektedir.

Araştırmanın ikinci boyutunda bireyin verdiği yanıtların doğruluğuna dair farkındalık düzeyini işaret eden üstbilişsel izleme becerisi ile matematiksel yeterlikler arasındaki ilişki belirlenmiştir. Üstbiliş, pek çok çalışmada üst düzey düşünme becerileri (higher-order thinking abilities) içinde yer almaktadır (ör., Brookhart, 2010; Conklin, 2012; Schraw ve Robinson, 2011; Williams, 2003). Araştırmada belirlenen dört beceri arasında ise, sadece “akıl yürütme ve strateji geliştirme” becerisine sahip olanlar ile bu beceriye sahip olmayan öğrencilerin üstbilişsel izleme performansları anlamlı olarak farklılaşmaktadır. Benzer bulgular, önceki çalışmalarda da gözlenmektedir (bkz., Kramarski ve Mevarech, 2003; Schneider ve Ardel, 2010). Örneğin, Kramarski ve Mevarech (2003), 8. sınıf öğrencileriyle yaptıkları araştırmalarında, birlikte öğrenme (cooperating learning) ile üstbilişsel eğitim verilen ve bireysel öğrenme (individual learning) ile yine üstbilişsel eğitim verilen grupların, sırasıyla yalnızca birlikte öğrenme ve yalnızca bireysel eğitim verilen gruplara göre, grafik oluşturma ve üstbilişsel bilgi performanslarında anlamlı derecede daha başarılı olduklarını göstermişlerdir.

Akıl yürütme ve strateji geliştirme becerisinin, üstbilişsel izleme performansı ile anlamlı derecede ilişkili bulunması alanyazında yer alan ve son yıllarda giderek daha fazla araştırmacı tarafından kullanılan örneğin, Koriat ve Goldsmith’in (1996) önerdiği modelle de oldukça uyumlu görülmektedir. “Bellek kesinliğinin stratejik olarak düzenlenmesi” (strategical regulation of memory accuracy) olarak adlandırılan bu çerçeve modeline göre, Koriat ve Goldsmith şunu önermektedir. Bir testte bulunan sorulara verilecek yanıtlar için, eğer serbest raporlama (free-report) yöntemi kullanılırsa (yani katılımcılardan bildikleri “yalnızca bütün doğru bilgileri” raporlamaları istenirse), katılımcılar, verdikleri yanıtların yalnızca doğru yanıtlar olmasını sağlamak amacıyla üstbilişsel bir strateji geliştirebilirler. Bu strateji için, örneğin, 20 sözcüğün bulunduğu bir sözcük listesini çalıştıktan sonra serbest-raporlama işlemine alınan bir katılımcı düşünelim ve “A katılımcısı” olarak adlandırabileceğimiz bu katılımcı yalnızca 12 sözcük hatırlayıp raporlamış olsun. Doğru olarak hatırladığını düşünerek raporladığı 12 sözcük arasında da sadece 8 tanesi gerçekten listede bulunan (doğru) sözcük olsun. “B” katılımcısının da yine aynı test için, 8 sözcük raporlamış olduğunu ve bu 8 sözcükten yalnızca 6’sını doğru olarak raporladığını düşünelim. Bu durumda, Koriat ve Goldsmith’e göre, B katılımcısı, A katılımcısına göre daha az “nicelik performansı (quantity performance)” sergilemiş olsa da (diğer bir ifadeyle, toplam çalışılan sözcük sayısına oranla (20) daha az doğru yanıt vermiş olsa da (6)), doğrularına ait “kesinlik” (accuracy) düzeyini stratejik olarak artırmıştır. Yani B katılımcısı, daha az yanıt vererek, yanıtlarındaki kesinlik derecesini yükseltebilmiştir. Modele göre rahatlıkla hesaplanabilen iki performanstan biri olan “nicelik performansı” raporlanan toplam doğru sayısının bütün olası doğru sayısına (örnek için, listedeki toplam sözcük sayısına) oranıdır. Bu örnekteki nicelik performansı, A katılımcısı için .40 (8/20) ve B katılımcısı için .30’dur (6/20). Raporlanan toplam doğru sayısının raporlanan toplam yanıt oranı olarak hesaplan “kesinlik performansı” ise, A katılımcısı için .66 (8/12) iken B katılımcısı için .75’tir (6/8). Özetle, bir katılımcı (katılımcı B) serbest-raporlamada daha az yanıt vermeyi seçerek ve bunu raporlama kriterini yüksek koyarak (aklına gelen doğru veya yanlış her sözcüğü rapor etmeyerek) ve/veya doğru ve yanlışlarını diğer katılımcı kadar iyi ayırt edemediğini fark ederek üstbilişsel bir strateji geliştirmiştir. Böylece, daha az sayıda raporlama yapmasına karşın “raporladıkları arasında daha fazla doğru yanıt vererek” bellek kesinliğini düzenleyebilmiştir. Tam da bu becerinin stratejik bir düzenleme olduğu ve başlı başına üst düzey bir bellek performansı olduğunu öneren Koriat ve Goldsmith’in modeline de paralel olarak, yürütülen bu çalışmada “akıl yürütme ve strateji geliştirme” becerisinin üstbilişsel izleme performansı ile yüksek derecede ilişkili bulunması kritik görünmektedir. Çalışma bulguları, Koriat ve Goldsmith’in modelinin temel varsayımından da hareketle, iletişim ve ilişkilendirme, matematikleştirme veya sembolik ve teknik dil kullanma gibi becerilerden ziyade, akıl yürütme ve strateji geliştirme becerisinin doğru ve yanlışları ayırt edebilme becerisiyle anlamlı olarak ilişkili olduğunu göstermiştir. Bu anlamda, yürütülen çalışmada STK’ya göre hesaplanan izleme performansının, BTM’den elde edilen becerilerin hangi veya hangileriyle ilişkili olabileceği, Koriat ve Goldsmith’in bellek kesinliğinin stratejik olarak düzenlenmesi modeliyle de yüksek düzeyde uyumlu olduğu görünmektedir.

Higham (2002) ve Koriat ve Goldsmith (1996) arasında, üstbilişsel performansların hangi hesaplamalarla daha iyi ölçülebileceğiyle ilgili çeşitli karşılaştırma çalışmalarının bulunmasına rağmen (bkz., Higham, 2011), bu çalışmanın bulguları, aslında üstbilişsel izleme performansı ve stratejik düzenlemenin birbirleriyle doğrudan ilişkili sayılabilecek beceriler olduğunu işaret etmektedir. “Düzenleme kesinliği” becerisinin Higham’ın önerdiği tip-2 STK yöntemiyle mi yoksa Koriat ve Goldsmith’in önerdiği yöntemle mi daha kesin şekilde hesaplanabileceğine dair çeşitli tartışmaların bulunmasına rağmen, aslında her iki yöntem de temelde aynı beceriyi ölçmeye çalışmaktadır. Yürütülen bu çalışmanın özgün değeri ise, izleme performansını Higham’ın (2002) önerdiği Tip-2 STK ile ölçerken, bu metotlardan bağımsız olan örtük sınıf analiziyle (ör., BTM ile) soru içeriğinin ölçtüğü beceriyi belirleyerek de (ör., “akıl yürütme ve strateji geliştirme” becerisi ile) izleme becerisine ait olabilecek olası örtük becerileri saptayabilmesi olarak görünmektedir. Ancak, buradaki temel soru, tersi olarak, neden yüksek akıl yürütme ve strateji geliştirme becerisinin iyi bir izleme becerisine sahip olmaktan kaynaklanmadığını sormak olabilir. Bu olası soruya yönelik olarak, bulgularımız ve çıkarımlarımız temelinde, üstbilişsel izleme becerisinin farklı alt becerilerden oluştuğunu düşünmekteyiz. Doğru ve yanlış yanıtları ayırt edebilmek, ancak akıl yürütme yapabilmek ve/veya belli bir bellek stratejisi geliştirebildikten sonra elde edilebilecek bir beceri olarak görünmektedir. Bu düşüncenin ilk nedeni kısaca, üstbilişsel izlemenin alanyazında bütüncül bir beceri olarak tanımlanmasıdır. Ek olarak, neden bazı kişilerin diğerlerine göre bu beceride daha iyi olduklarının nedeni de tam olarak belirlenmiş görünmemektedir. Özellikle yaşa bağlı bilişsel becerilerin değişimini inceleyen pek çok çalışma, hipotez testi kullanan bu çalışmadan farklı olarak, çoğunlukla korelasyon çalışmaları olarak karşımıza çıkmaktadır. Bu soruya farklı bir deney deseniyle yanıt verebilmek için, gelecekteki çalışmaların, örneğin, strateji geliştirme becerisi kazandırılan öğrencilerin üstbilişsel izleme performanslarını kontrol grubu ile karşılaştırarak, strateji geliştirme izleme performansı üzerindeki doğrudan etkisine destek bulabilecekleri düşünülmektedir. Ayrıca, devam eden çalışmalarda yine izleme performansı ile ilgili diğer olası alt becerilerin saptanmasının ve yine izleme becerisinin PISA ya da diğer ilgili testlerin alt bölümlerinde (ör., fen bilgisi, dil becerisi gibi) ve farklı yaş gruplarında da incelenmesinin kritik olduğu düşünülmektedir.

Yukarıda bahsedilen önerilerin yanında, araştırmanın bir takım sınırlılıkları da bulunmaktadır. Araştırma öncelikle 6 sınıf düzeyindeki matematik alanındaki üst düzey düşünme becerilerini ölçmeye yönelik bir test üzerinden planlanmıştır. Bu nedenle, daha temel özellikler için geliştirilmiş bir testte benzer bir çalışmanın yapılması bu çalışmada ulaşılan bulguların doğrulanması açısından önemli görülmektedir. Aynı zamanda özellikle BTM modelleri için farklı yaklaşımların karşılaştırılması, alanyazının geliştirilmesi açısından da önemli olduğu düşünülmektedir. Gerek bilişsel psikoloji, gerekse psikometri açısından güncel olan iki modelin aynı anda incelendiği bir araştırma için örneklem büyüklüğünün artırılması, test uzunluğunun yeniden düzenlenmesi, konu alanı ve sınıf düzeyi gibi değişkenlerin farklılaştırılmasıyla, bu konu ile ilgili alanyazına önemli katkı sağlayacağı söylenebilir.

Özetle, araştırma sonuçları BTM’ye göre kurulan sınıflama modelinin geçerliğine ilişkin bir çapraz-geçerlik (cross-validation) kanıtı sunabilmektedir. Özellikle örtük sınıf temelli çalışmalarda kurulan modellerin geçerliğine ilişkin sadece istatistiksel değil, aynı zamanda mantıksal kanıtların da araştırılması gerektiği düşünülmektedir. Bu çalışmanın bulguları, BTM sınıflamalarının mantıksal geçerliğine ilişkin görgül sonuçlar sunan ender araştırmalardan biri olarak karşımıza çıkmaktadır. Böylece, araştırmanın gelecekteki çalışmalar için önemli bir referans kaynağı olabilme potansiyeli taşıyabileceği düşünülmektedir. Özellikle BTM modellerinin model ve veri uyumu ve ölçüm geçerliği çalışmalarının bağlı ölçütlere dayanması ve dolayısıyla mutlak bir kritere göre değerlendirilemeyen sonuçlar ürettiği göz önüne alınırsa, model veri uyumu için psikolojik yapı ile psikometrik bulguların birbirini desteklemesi, modellerin kullanım alanlarının genişlemesi açısından da önemli bir araştırma konusu oluşturabileceği önerilmektedir.

Kaynakça

- Abdi, H. (2007). Signal detection theory. N. Salkind (Ed.), *Encyclopedia of measurement and statistics*. New York: Elsevier.
- Albacete, R. O., Cascón, J. A., Arnal, L. A., Pérez, J. D., Domínguez, I. E. ve Sánchez, F. S. (2016). Psychometric properties of the reading comprehension test ECOMPLEC. *Sec. Psicothema*, 28(1), 89-95. doi:10.7334/psicothema2015.92
- Ardelt, C. ve Schneider, W. (2015). Cross-country generalizability of the role metacognitive knowledge for students' strategy use and reading competence. *Teachers College Record*, 117(1), 1-32.
- Ardelt, C., Schiefele, U. ve Schneider, W. (2001). Predictors of reading literacy. *European Journal of Education*, 16, 363-383. doi:10.1007/BF03173188
- Bahrnick, H. P. (1970). Two-phase model for prompted recall. *Psychological Review*, 77, 215-222.
- Baltes, P. B., Staudinger, U. M. ve Lindenberger, U. (1999). Lifespan psychology: Theory and application to intellectual functioning. *Annual Review of Psychology*, 50(1), 471-507.
- Başokçu, T. O. (2012). DINA model parametreleri kullanılarak tahminlenen madde ayırıcılık indekslerinin incelenmesi. *Eğitim ve Bilim*, 37(163), 310-321.
- Başokçu, T. O. (2014). Öğrenci yeteneklerinin kestirilmesinde bilişsel tanı modelleri ve uygulamaları. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 14(1), 1-32. doi:10.17240/aibuefd.2014.14.1-5000091500
- Başokçu, T. O. (2019). *A recommended model to increase success level of turkey in mathematics in international wide scale exams. Effectiveness of the cognitive diagnosis based tracking model*. Izmir: TUBITAK 115K531.
- Bean, J. C. ve Peterson, D. (1998). Grading classroom participation. *New Directions for Teaching & Learning*, 74, 33-40. doi:10.1002/tl.7403
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Danver, MA: ASCD Publications. <https://books.google.com.tr/books?id=AFIxeGsV6SMC> adresinden erişildi.
- Chen, J., de la Torre, J. ve Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement*, 50(2), 123-140. doi:10.1111/j.1745-3984.2012.00185.x
- Conklin, W. (2012). *Strategies for developing higher-order thinking skills*. CA: Shell Education.
- de la Torre, J. (2008). An empirically based method of q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, 45(4), 343-362. doi:10.1111/j.1745-3984.2008.00069.x
- de la Torre, J. ve Chiu, C.-Y. (2015). A general method of empirical q-matrix validation. *Psychometrika*, 81, 1-21. doi:10.1007/s11336-015-9467-8
- de la Torre, J. ve Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69(3), 333-353.
- de la Torre, J. ve Lee, Y.-S. (2010). A note on the invariance of the DINA model parameters. *Journal of Educational Measurement*, 47(1), 115-127. doi:10.1111/j.1745-3984.2009.00102.x
- DiBello, L. V., Roussos, L. A. ve Stout, W. (2006). A review of cognitively diagnostic assessment and a summary of psychometric models. C. R. Rao ve S. Sinharay (Ed.), *Handbook of statistics içinde* (26. cilt, s. 979-1030). Hollanda: Elsevier. doi:10.1016/S0169-7161(06)26031-0
- Dunlosky, J. ve Tauber, S. (2001). *The Oxford handbook of metamemory*. NY: Oxford University Press.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist*, 34(10), 906-911. doi:10.1037/0003-066X.34.10.906
- George, A., Robitzsch, A., Kiefer, T., Groß, J. ve Ünlü, A. (2016). The R package CDM for cognitive diagnosis models. *Journal of Statistical Software*, 74(2), 1-24. doi:10.18637/jss.v074.i02

- Gierl, M. J., Alves, C. ve Majeau, R. T. (2010). Using the attribute hierarchy method to make diagnostic inferences about examinees' knowledge and skills in mathematics: An operational implementation of cognitive diagnostic assessment. *International Journal of Testing*, 10(4), 318-341. doi:10.1080/15305058.2010.509554
- Green, D. M. ve Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Güzel, M. A. ve Higham, P. A. (2013). Dissociating early- and late-selection processes in recall: The mixed blessing of categorized study lists. *Memory & Cognition*, 41, 683-697. doi:10.3758/s13421-012-0292-3
- Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, 26(4), 301-321. doi:10.1111/j.1745-3984.1989.tb00336.x
- Handel, M., Ardelt, C. ve Weinert, S. (2013). Assessing metacognitive knowledge: Development and evaluation of a test instrument. *Journal for Educational Research Online*, 5(2), 162-188. https://www.pedocs.de/frontdoor.php?source_opus=5481 adresinden erişildi.
- Henson, R. ve Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement*, 29(4), 262-277. doi:10.1177/0146621604272623
- Higham, P. A. (2002). Strong cues are not necessarily weak: Thomson and Tulving (1970) and the encoding specificity principle revisited. *Memory & Cognition*, 30(1), 67-80. doi:10.3758/BF03195266
- Higham, P. A. (2007). No special K! A signal detection framework for the strategic regulation of memory accuracy. *Journal of Experimental Psychology: General*, 136, 1-22. doi:10.1037/0096-3445.136.1.1
- Higham, P. A. (2011). Accuracy discrimination and type-2 signal detection theory: Clarifications, extensions, and an analysis of bias. P. A. Higham ve J. P. Leboe (Ed.), *Constructions of remembering and metacognition. Essays in honour of Bruce Whittlesea* içinde (s. 109-127). Basingstoke: Palgrave Macmillan. doi:10.1057/9780230305281_9
- Higham, P. A. ve Arnold, M. M. (2007). How many questions should I answer? Using bias profiles to estimate optimal bias and maximum score on formula-scored tests. (In special issue on: Bridging cognitive science and education: learning, memory, and metacognition.). *European Journal of Cognitive Psychology*, 19, 718-742. doi:10.1080/09541440701326121
- Higham, P. A. ve Gerrard, C. (2005). Not all errors are created equal: metacognition and changing answers on multiple-choice tests. *Canadian Journal of Experimental Psychology*, 59, 28-34. doi:10.1037/h0087457
- Higham, P. A. ve Tam, H. (2005). Generation failure: Estimating metacognition in cued recall. *Journal of Memory and Language*, 52(4), 595-617. doi:10.1016/j.jml.2005.01.015
- Hu, J., Miller, M. D., Huggins-Manley, A. C. ve Chen, Y.-H. (2016). Evaluation of model fit in cognitive diagnosis models. *International Journal of Testing*, 16(2), 119-141. doi:10.1080/15305058.2015.1133627
- Huo, Y. ve de la Torre, J. (2014). Estimating a cognitive diagnostic model for multiple strategies via the EM algorithm. *Applied Psychological Measurement*, 38(6), 464-485. doi:10.1177/0146621614533986
- Junker, B. W. (1999). Some statistical models and computational methods that may be useful for cognitively-relevant assessment. *Committee on the Foundations of Assessment*. <http://www.stat.cmu.edu/~brian/nrc/cfa/documents/final.pdf> adresinden erişildi.
- Junker, B. W. ve Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with non-parametric item response theory. *Applied Psychological Measurement*, 25, 258-272. doi:10.1177/01466210122032064
- Karakelle, S. ve Saraç, S. (2010). Üst biliş hakkında bir gözden geçirme: Üstbiliş çalışmalarını mı yoksa üst bilişsel yaklaşımı mı?. *Türk Psikoloji Yazıları*, 13(26), 45-60.
- Kintsch, W. (1970). Models for free recall and recognition. D. A. Norman (Ed.), *Models of human memory* içinde (s. 333-370). New York: Academic Press.
- Koriat, A. ve Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103(3), 490-517. doi:10.1037/0033-295X.103.3.49

- Kramarski, B. ve Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal*, 40(1), 281-310. doi:10.3102/00028312040001281
- Leighton, J. ve Gierl, M. (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge: Cambridge University Press.
- Lie, S., Taylor, A. ve Harmon, M. (1996). Scoring techniques and criteria. M. O. Martin ve D. L. Kelly (Ed.), *Third International Mathematics and Science Study (TIMSS) technical report. Volume I: Design and development*. Chestnut Hill, MA: Boston College.
- Lindblom-Ylänne, S., Pihlajamäki, H. ve Kotkas, T. (2006). Self, peer, and teacher assessment of student essays. *Active Learning in Higher Education*, 7(1), 51-62. doi:10.1177/1469787406061148
- Luce, R. D. ve Krumhansl, C. L. (1988). Measurement, scaling, and psychophysics. R. C. Atkinson, R. J. Herrnstein, G. Lindzey ve R. D. Luce (Ed.), *Stevens' handbook of experimental psychology* içinde (Cilt 1, 2. bs., s. 3-74). New York: Wiley.
- Maag Merki, K., Ramseier, E. ve Karlen, Y. (2013). Reliability and validity analyses of a newly developed test to assess learning strategy knowledge. *Journal of Cognitive Education and Psychology*, 12(3), 391-408. doi:10.1891/1945-8959.12.3.391
- MacReady, G. B. ve Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, 2(2), 99-120. doi:10.2307/1164802
- Myers, M. ve Paris, S. G. (1978). Children's metacognitive knowledge about reading. *Journal of Educational Psychology*, 70(5), 680-690. doi:10.1037/0022-0663.70.5.680
- OECD. (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD. doi:10.1787/9789264101739-en
- OECD. (2010). *PISA 2009 results: What students know and can do. Student performance in reading, mathematics and science* (1. cilt). Paris: OECD. doi:10.1787/9789264091450-en
- OECD. (2019). *PISA 2018 assessment and analytical framework*. Paris: OECD Publishing. doi:10.1787/b25efab8-en
- Pieschel, S. (2009). Metacognitive calibration - an extended conceptualization and potential applications. *Metacognition and Learning*, 4(1), 3-31. doi:10.1007/s11409-008-9030-4
- Schneider, W. ve Ardet, C. (2010). Metacognition and mathematics education. *ZDM*, 42(2), 149-161. doi:10.1007/s11858-010-0240-2
- Schraw, G. ve Robinson, D. H. (2011). *Assessment of higher order thinking skills*. Kuzey Carolina: Information Age Pub. <https://books.google.com.tr/books?id=6wAoDwAAQBAJ> adresinden erişildi.
- Stacey, K. ve Turner, R. (2014). *Assessing mathematical literacy: The PISA experience*. Cham: Springer International Publishing.
- von Davier, M. (2014). The DINA model as a constrained general diagnostic model: Two variants of a model equivalency. *British Journal of Mathematical and Statistical Psychology*, 67(1), 49-71. doi:10.1111/bmsp.12003
- Watkins, M. J. ve Gardiner, J. M. (1979). An appreciation of generate-recognize theory of recall. *Journal of Verbal Learning & Verbal Behavior*, 18(6), 687-704. doi:10.1016/S0022-5371(79)90397-9
- White, B. ve Frederiksen, J. (2005). A theoretical framework and approach for fostering metacognitive development. *Educational Psychologist*, 40, 211-223. doi:10.1207/s15326985ep4004_3
- Williams, R. B. (2003). *Higher order thinking skills: Challenging all students to achieve*. CA: SAGE Publications.
- Wirth, J. ve Leutner, L. (2008). Self regulated learning as a competence. Implication of theoretical models for assessment methods. *Journal of Psychology*, 216(2), 102-110. doi:10.1027/0044-3409.216.2.102
- Wragg, E. C. (2001). *Questioning in the secondary school*. GBR London: Routledge.