
Sato Test Kuramı ile Yapılan Sınıflamalara Yönelik Öğretmen Görüşlerinin İncelenmesi ¹

Sait ÇÜM², Elif Kübra DEMİR³, Murat Doğan ŞAHİN⁴

Geliş Tarihi: 19.12.2017

Kabul Tarihi: 20.01.2018

Araştırma Makalesi

Öz

Bu araştırmanın amacı, Sato Test Kuramı (STK)'na dayalı olarak gerçekleştirilen ve maddelerin işlerliği ile öğrencilerin öğrenme durumları hakkında dönüt sağlayan sınıflamalara yönelik öğretmen görüşlerinin toplanması ve incelenmesidir. Bu amaçla MEB'de farklı düzeylerde görev yapan 70 öğretmene çevrimiçi ve yüzyüze olmak üzere 12 maddelik bir anket uygulanmıştır. Betimsel bir yöntem izlenerek öğretmenlerin her bir anket maddesine verdiği yanıtların frekans değerleri ve yüzdeleri raporlanmıştır. Elde edilen sonuçlar, katılımcı öğretmenlerin büyük çoğunluğunun (%81.4) STK'yı kullanmaya istekli olduklarını göstermektedir. Bununla birlikte katılımcıların, STK ile test maddelerinin işlerliği ve öğrencilerin öğrenme durumları hakkında verilen dönütleri beğendikleri ve işe yarar buldukları belirlenmiştir. Ayrıca öğretmenlerin büyük çoğunluğu dönütleri madde havuzu oluşturma konusunda teşvik edici bulmakta ve öğretimi iyileştirme noktasında işlerini kolaylaştıracağını düşünmektedirler. Buna karşın, bu araştırmaya katılan öğretmenlerin önemli bir kısmının maddeler ve öğrencilerle ilgili dönütlerin yeterli olup olmadığına ilişkin kararsızlık yaşadıkları görülmüştür. Ayrıca katılımcıların %41.4'ü madde analizleri için klasik yöntemleri tercih edip etmeme noktasında kararsızlık yaşadığını ifade etmiştir.

Anahtar kelimeler: Sato test kuramı, öğrenci-madde çizelgesi analizi, madde ve test analizi

¹ Bu çalışma 22-23 Aralık 2017 tarihleri arasında İstanbul'da düzenlenen I. Uluslararası Sosyal Beşeri ve Eğitim Bilimleri Kongresi'nde sözlü bildiri olarak sunulmuştur.

² Milli Eğitim Bakanlığı, e-mail: saitcum@hotmail.com

³ Ege Üniversitesi, e-mail: elifkubrademir@gmail.com

⁴ Anadolu Üniversitesi, e-mail: mdsahin@anadolu.edu.tr

Examination of Teacher Opinions on Classification by Sato Test Theory

Submitted by 19.12.2017

Accepted by 20.01.2018

Research Paper

Abstract

The aim of this study is to collect and examine teachers' opinions about classifications based on the Sato Test Theory (STT) and provide feedback on the functioning of the items and the learning situations of the students. For this purpose, a 12-item questionnaire was applied to 70 teachers who were working at different levels in Turkish National Education Ministry schools, online and face-to-face. By following a descriptive method, frequency values and percentages of the responses given by the teachers to each item of the questionnaire were reported. The results show that the vast majority of participating teachers (81.4%) are willing to use the STT. However, it was determined that the participants were satisfied with the feedback provided by STT about the functioning of the test items and the learning situations of the students. In addition, the vast majority of teachers believe that the feedbacks of STT encourages them to build an item pool and facilitates their work at the point of improving teaching. However, it was observed that a significant proportion of the teachers participated in this study had an ambiguity about whether the feedbacks related to the items and the students were sufficient. It is also a remarkable finding that 41.4% of the participants expressed their hesitancy about choosing classical methods for substance analysis.

Keywords: Sato test theory, student-problem chart analysis, item and test analysis

Giriş

Eğitimde ve psikolojide test geliştirme ve ölçme süreçleri, oldukça karmaşık ve geniş kapsamlı bilgiler bütünü üzerine inşa edilmiş bir bilim dalı olan psikometri çerçevesinde ele alınır. Psikometristler, test maddelerinin nasıl geliştirileceği, nasıl seçileceği, verilen cevapların sayılara nasıl dönüştürüleceği, bunların nasıl yorumlanacağı ve sonuçların nasıl değerlendirileceği gibi sorulara matematiksel ve istatistiksel kuramlara dayalı olarak yanıt ararlar. Bu kuramlar, “test kuramı” ya da “psikometrik kuram” şeklinde adlandırılmaktadır (Suen, 1990).

Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı (MTK) geçmişten günümüze üzerinde en fazla çalışılan başlıca iki test kuramı olarak öne çıkmaktadır. KTK, gözlenen ile gözlenemeyen özellikler arasındaki ilişkiyi doğrusal bir model aracılığıyla açıklar (Lord ve Novick, 1968). Kuramda, ölçülmek istenen değişkene ait gerçek değer “gerçek puan” olarak adlandırılır ve kurama göre gerçek puan sabit olsa da gözlem sonuçları (gözlenen puan) pratikte her zaman bir miktar hata içerir. Bu üç bileşen arasındaki eşitlik $X = T + E$ şeklinde kurulur. Söz konusu eşitlikte “X” gözlenen puanı, “T” gerçek puanı ve “E” ölçmeye karışan hataları temsil etmektedir (Crocker ve Algina, 1986; Hambleton ve Swaminathan, 1985; Thompson, 1994). Özellikle, okullarda uygulanan başarı testlerini geliştiren öğretmenler ve araştırmaları kapsamında çeşitli ölçekleri kullanarak bilgi toplayan fakat ölçme alanında üst düzey bilgi sahibi olmayan araştırmacılar için KTK, madde ve test istatistiklerinin hesaplanması bakımından daha tercih edilir bir kuram olarak geçerliğini korumaktadır. Buna karşın, KTK’yı tartışma konusu haline getiren bazı dezavantajlar bulunmaktadır. Söz konusu dezavantajlar, maddeler halinde aşağıda özetlenmiştir.

- KTK'ya dayalı geliştirilen testler/ölçeklerde madde istatistikleri testin geliştirdiği gruptan kestirilmekte ve grup değiştiğinde elde edilen istatistik değerleri de değişebilmektedir.
- KTK'da test puanları da madde örnekleme bağılıdır. Eğer test göreceli olarak kolay maddelerden oluşuyorsa bireylerin testten aldığı puanlar da yüksek olacak; test göreceli olarak zor maddelerden oluşuyorsa bireylerin testten aldığı puanlar da düşük olacaktır. Bu bağlamda, KTK'da bireylerin başarı/yetenek düzeylerinin de testte yer alan maddelere bağlı olarak göreceli olduğu söylenebilir (McDonald ve Paunonen, 2003).
- KTK modellemesinde tüm hatalar tesadüfi hata olarak kabul edilmektedir. Buna karşın kültürel altyapı, etnisite, cinsiyet gibi kaynaklardan ileri gelen sistematik hatalar ölçmeye karışabilir. Test yanlılığına neden olan bu tür hatalar KTK'da görmezden gelinir (Kiany ve Jalali, 2009).
- KTK'da ölçümlerin güvenilirliği testi yanıtlayan bireylerin özelliklerine, testin özelliklerine ve test uygulayıcısının kararlarına ve yönetimine bağlı olarak değişebilir (Traub ve Rowley, 1991).
- KTK'ya dayalı uygulamalarda teste katılan bireylere verilen en önemli bilgi, onun tüm test üzerinde gösterdiği performansını yansıtan toplam test puanıdır. Bireyin farklı maddeler üzerinde gösterdiği performans KTK'da dikkate alınmaz (Kiany ve Jalali, 2009).

KTK'nın sınırlılıklarını ortadan kaldıracabileceği iddia edilen bir model olarak Madde Tepki Kuramı (MTK) 1930'lu yılların sonunda ortaya atılmıştır. Tucker 1940'da kuramın en önemli kavramlarından "madde karakteristik eğrisi" kavramını ilk defa kullanmıştır. Lord'un 1953'te yayımlanan makalesi ve Birnbaum'un 1957 ve 1958 yılları içerisinde Amerika Birleşik Devletleri Hava Kuvvetleri için yazdığı üç teknik rapor MTK'ya ilişkin çalışmaların

başlangıcı olarak kabul edilir. Lord ve Novick'in (1968) "Statistical Theories of Mental Test Scores" kitabında Birnbaum tarafından MTK üzerine kaleme alınan dört bölüm kuramın gelişmesinde büyük rol oynamıştır. 1950-1970 yılları arasında matematiksel işlemlerin karmaşıklığı nedeniyle kuram yavaş gelişmiş fakat bilgisayar teknolojisindeki gelişmelere paralel olarak kuramın kullanımı ve gelişimi de hızlanmıştır (Hambleton ve Swaminathan, 1985; Embretson ve Reise, 2000).

MTK'nın tek boyutluluk, yerel bağımsızlık ve maddelere verilen yanıtlarla bireylerin yetenekleri arasındaki ilişkilerin madde karakteristik eğrisi fonksiyonlarıyla ifade edilebileceği gibi varsayımları bulunmaktadır. Bu varsayımlar ve model-veri uyumu sağlandığı takdirde MTK'yı KTK'dan üstün kılan bazı avantajların olduğu belirtilmektedir (Doğan ve Tezbaşaran, 2003; Fan, 1998; Hambleton ve Swaminathan, 1985). Söz konusu avantajların büyük çoğunluğu MTK'da madde parametrelerinin testin uygulandığı örneklemden bağımsız yetenek parametresinin de bireyin aldığı maddelerden bağımsız olarak kestirilebilmesinin sonucu olarak ortaya çıkar. Pratikte bu durum test eşitleme ve bireye uyarlanmış test geliştirme gibi alanlarda kolaylık sağlar. Ayrıca madde özelliklerinin örneklemden bağımsız oluşu bir testin MTK'ya göre bir kez ölçeklendikten sonra farklı örneklemlerde pek çok kez kullanılabilmesine ve elde edilen sonuçların karşılaştırılabilmesine olanak sağlar (Hambleton ve Swaminathan, 1985; Hambleton, Swaminathan ve Rogers, 1991). Diğer taraftan, MTK'ya göre hesaplanan yetenek puanlarına her bir maddenin katkısı aynı değildir; kullanılan modele göre maddenin güçlüğünden, ayırt edicilik gücünden ve tahminle doğru yanıtlanma olasılığından etkilenen yetenek puanları söz konusudur (Çelen ve Aybek, 2013). Bu sayede MTK'nın, KTK'ya dayalı ölçmelerde sık karşılaşılan eşit sayıda fakat farklı maddelere doğru yanıt vermiş olan bireylerin birbirinden ayırt edilememesi sorununu ortadan kaldırdığı yorumu yapılabilir. Embretson ve Reise'a göre (2000) MTK'nın

bir başka avantajı da KTK'nın daha uzun testler daha güvenilirdir kabulünün aksine daha az sayıda maddeden oluşan, güvenilirliği yüksek testler geliştirilebilmesine olanak sağlamasıdır.

Sözü edilen avantajlar MTK varsayımları ve model-veri uyumu sağlanamadığı takdirde geçerliğini kaybetmektedir. Ayrıca MTK parametrelerini doğru bir şekilde kestirebilmek için büyük örneklemeler gerekmektedir (de Ayala, 2009; DeMars, 2016; Drasgow, 1982; Hulin, Lissak, Ree ve Jensen, 1983; van der Linden ve Hambleton, 1997). Bununla birlikte, MTK'ya dayalı bir test geliştirmenin, uygulamalardan elde edilen verileri analiz etmenin ve analiz sonuçlarını yorumlamanın alan uzmanlığı gerektirmesi gibi durumlardan kaynaklı olarak kuramın, sınıf içi ölçme ve değerlendirme uygulamalarında kullanılmasının önünde aşılması zor engellerin olduğu görüşü ileri sürülebilir.

Takahiro Sato tarafından 1969'da ileri sürülen Sato Test Kuramı (STK) ise öğrencilerin cevap örüntüleri üzerinden öğrenme durumlarını belirlemeye yönelik olarak normal dağılım varsayımını sağlayamayan küçük örneklemelerde bile etkili sonuçlar üretmesi beklenen analiz yöntemleri üzerine inşa edilmiştir. STK, ilk başlarda yalnızca Öğrenci Madde Çizelgesi Analizi (Student-Problem Chart Analysis) ile öğrenme durumları bakımından öğrencileri ve kaliteleri bakımından maddeleri sınıflandırma amaçlarına yönelik olarak kullanılmış olsa da 2010 yılında Nagai'nın önerisiyle Ju-Long Deng tarafından ileri sürülen Gri İlişki Analizi ile geliştirilmiş ve öğrencilerin başarı/yetenek düzeylerinin belirlenmesi amacıyla da kullanılabilir hale getirilmiştir. Günümüzde STK, Gri Öğrenci-Madde Çizelgesi Analizi (Grey S-P Chart Analysis), Kısmi Kredili Öğrenci-Madde Çizelgesi Analizi (Partial Credit S-P Chart Analysis) ve Rasch Gri Öğrenci-Madde Çizelgesi Analizi (Rasch Grey S-P Chart Analysis) modelleri ile hem ikili puanlanan hem de çoklu puanlanan madde türlerine uygulanabilen, test geliştirme ve uygulama alanında mevcut sorunlara çözüm getirebilecek yöntemler sunma çabasıdadır (Sheu ve vd., 2014a; Pham, Sheu ve Nagai, 2015).

Öğrenci-Madde Çizelgesi Analizi ile *Öğrenci Uyarı İndeksi* (Student Caution Index) ve *Madde Uyarı İndeksi* (Problem Caution Index) olmak üzere iki tür indeks üretilmektedir.

Madde uyarı indeksi (MUI) aracılığıyla maddeler sınıflandırılarak her bir maddenin işlevliliği hakkında dönüt elde edilmektedir. MUI indeksi aşağıdaki formülle hesaplanmaktadır (Lin ve Yih, 2015; Sheu vd., 2014b).

$$MUI = 1 - \frac{\sum_{n=1}^N (Ynm)(Yn) - (Ym) \cdot (\mu')}{\sum_{n=1}^N Yn - (Ym) \cdot (\mu')}$$

m: Maddeler ($m=1,2,3,\dots,M$)

n: Öğrenciler ($n=1,2,3,\dots,N$)

Ynm: *n*. öğrencinin *m*. maddeden aldığı puan (doğru: 1, yanlış: 0)

Ym: Öğrencilerin *m*. maddeden aldıkları puanların toplamı

Yn: *n*. öğrencinin toplam test puanı

$$\mu' = \frac{\sum_{n=1}^N Yn}{N}$$

Madde uyarı indeksi (MUI), maddeleri dört farklı sınıfa atamak için kullanılmaktadır. Her bir madde için hesaplanan MUI değerinin 0.50'den daha büyük bir değer alması ilgili maddede bazı sorunların olabileceğine ilişkin bilgi verir. Elde edilen MUI değeri maddenin doğru yanıtlanma oranı ile birlikte yorumlanır. Yapılan sınıflandırmalar ve bunlara ilişkin yapılabilecek yorumlar aşağıda verilmiştir (Sheu, Pham, Nguyen ve Nguyen, 2013).

A: Düzgün çalışan bir madde.

A': Kısmi olarak sorunlu bir madde. Tekrar gözden geçirilebilir.

B: Zor bir madde.

B': Sorunlu bir madde. Ters işliyor olabilir. Testten atılmalı ya da yeniden yazılmalı.

Öğrenci uyarı indeksi (ÖÜİ) aracılığıyla öğrenciler sınıflandırılarak her bir öğrencinin testle ölçülen konu alanlarını öğrenme ve maddeleri dikkatli bir şekilde yanıtlama durumlarına ilişkin dönüt alınmaktadır. ÖÜİ indeksi aşağıdaki formülle hesaplanmaktadır (Lin ve Yih, 2015; Sheu vd., 2014b).

$$\text{ÖÜİ} = 1 - \frac{\sum_{m=1}^M (Y_{nm})(Y_m) - (Y_n) \cdot (u')}{\sum_{m=1}^M Y_m - (Y_n) \cdot (u')}$$

m: Maddeler ($m=1,2,3,\dots,M$)

n: Öğrenciler ($n=1,2,3,\dots,N$)

Y_{nm}: n. öğrencinin m. maddeden aldığı puan (doğru: 1, yanlış: 0)

Y_m: Öğrencilerin m. maddeden aldıkları puanların toplamı

Y_n: n. öğrencinin toplam test puanı

$$u' = \frac{\sum_{m=1}^M Y_m}{M}$$

Öğrenci uyarı indeksi (ÖÜİ), öğrencileri altı farklı sınıfa atamak için kullanılmaktadır. Her bir öğrenci için hesaplanan ÖÜİ değerinin 0.50'den daha büyük bir değer alması, ilgili öğrencinin öğrenme eksikliklerinin olabileceğine ya da dikkatsizlikle kaçırdığı soruların olabileceğine ilişkin bilgi verir. Elde edilen ÖÜİ değeri öğrencinin doğru yanıt verdiği madde sayısının testte yer alan madde sayısına oranı ile birlikte yorumlanır. Yapılan sınıflandırmalar ve bunlara ilişkin yapılabilecek yorumlar aşağıda verilmiştir (Sheu, Pham, Nguyen ve Nguyen, 2013).

A: Etkili öğrenme gerçekleşmiş.

A': Öğrenme var fakat çok dikkatsiz.

B: Genel olarak iyi ancak biraz daha fazla çalışmaya ihtiyacı var.

B': Biraz dikkatsiz ve daha fazla çalışmaya ihtiyacı var.

C: Öğrenme düzeyi yetersiz.

C': Öğrenme gerçekleşmemiş.

STK, hem ikili puanlanan hem de çoklu puanlanan maddeler için uygun modeller sunmaktadır. Bu bakımdan geniş bir kullanım alanının varlığından söz edilebilir. Yapılan çalışmalarda ulaşılan sonuçlar ve ilgili alanyazın doğrultusunda STK'nın avantajları şöyle özetlenebilir:

- Kurama dayalı olarak hesaplanan ÖÜİ ve MÜİ indeksleri ile kavram yanılgılarına sahip olan öğrencilerin ve kavram yanılgısı içeren maddelerin tespit edilebildiği ileri sürülmektedir (Tsai vd., 2014).
- STK'dan, düzey belirlemeye ve buna dayalı olarak öğrencilerin eğitim durumları ile karar almaya (özetleyici değerlendirme) yönelik süreçlerde yararlanılabildiği gibi, öğrencilerin öğrenme durumları ve öğretim sürecinin etkililiği ile ilgili dönüt alabilmek ve alınan dönütlere dayalı olarak öğretimi geliştirmek (biçimlendirici değerlendirme) amaçlarıyla da yararlanılabilmektedir (Switzer ve Connell, 1990; Yu, 2011).
- Öğrencilerle ilgili verilen nitel dönütler ve öğrencilerin öğrenme durumlarına göre sınıflandırılması, öğretmenlerin dersi bireyselleştirilmiş öğretime uygun olarak düzenlemesine yardımcı olabilir. Maddelere ilişkin verilen nitel dönütler ise öğretmenleri, kaliteli maddelerden oluşan bir madde havuzu oluşturma konusunda cesaretlendirebilir.
- STK, hem ikili puanlanan hem de çoklu puanlanan maddeler için uygun modeller sunmaktadır. Bu bakımdan geniş bir kullanım alanı vardır. İçerdiği analizler basit ve

kullanışlıdır. Analiz çıktılarının yorumu uzmanlık gerektirmez (McArthur, 1983; Pham, Sheu ve Nagai, 2015).

- Öğrenci-Problem Çizelgesi Analizi'nin yapılabilmesi için veriye (dataya) ilişkin herhangi bir varsayımın sağlanmasına ihtiyaç duyulmamaktadır. Bu bakımdan küçük gruplar üzerinde yürütülen ölçme ve değerlendirme uygulamaları için uygun ve kullanışlıdır (Pham, Sheu ve Nagai, 2015).
- Çüm, Gelbal ve Tsai (2016), farklı örneklemlerden (30'ar kişilik) alınan aynı maddelere ait yanıt örüntüleri üzerinden elde ettikleri MİSGİD indekslerinin tutarlı ve MUI indeksleri aracılığıyla yapılan madde sınıflandırmalarının uyumlu olduğu sonucuna ulaşmışlardır. Bu sonuç, KTK'nın, madde istatistiklerinin testin uygulandığı gruba bağlı olmasıyla ilgili olan dezavantajına karşı STK'nın sağladığı bir avantaj olarak öne sürülebilir.

STK, Japon ilk ve ortaokullarında önemli bir ölçme ve değerlendirme yaklaşımı olarak kabul görmektedir (Sato, 1984). Buna karşın Türkiye'de kurama henüz yeterince ilgi duyulmadığı ve hatta çok az sayıda çalışmada (Acar, 2006; Bayrak ve Yurdugül, 2016; Çüm, Gelbal ve Tsai, 2016) kendisinden söz edildiği belirtilebilir. Bir kuramın kullanımının yaygınlaşmasında bilim çevrelerince ve o kuramdan pratikte fayda sağlayacak kişilerce genel bir kabul görmesinin etkili olduğu görüşü ileri sürülebilir. STK'nın ortaya çıkış iddiasının ölçme ve değerlendirmenin en sık tekrarlandığı eğitim ortamlarında geçerli ve güvenilir sonuçlar elde etmek olduğu dikkate alındığında, yalnızca alan uzmanlarının ya da test/ölçek geliştiren araştırmacıların değil öğretmenlerin kurama yönelik yaklaşımlarının belirlenmesi de önemli görülmektedir. Yapılan tartışmalardan hareketle bu çalışmada, Sato Test Kuramı ile yapılan sınıflandırmalara yönelik öğretmen görüşlerinin toplanması ve incelenmesi amaçlanmaktadır. Bu tür araştırmaların kurama olan ilgiyi artıracak ve ardından yapılacak olan çalışmalara öncülük edeceği düşünülmektedir.

Yöntem

Araştırmada betimsel bir yöntem izlenerek öğretmenlerin Sato Test Kuramı ile yapılan sınıflamalara yönelik görüşleri alınmıştır. Bu bölümde çalışma grubu, kullanılan ölçme aracı ve elde edilen verilerin analizi ile ilgili detaylar paylaşılmıştır.

Çalışma Grubu

Çalışma grubu, gönüllülük esasına dayalı olarak araştırmaya katılan 70 öğretmenden oluşmaktadır. Araştırmacılar tarafından oluşturulan anket, katılımcıların 51'ine çevrimiçi olarak, kalan 19 kişiye ise yüz yüze uygulanmıştır. Özellikle çevrimiçi yanıtlayıcıların sayısının çok olması, araştırma grubunun heterojen bir yapıya sahip olmasını sağlamıştır. Katılımcı çokluğu bakımından başta Burdur, İstanbul, Kahramanmaraş, Kırıkkale, Kırşehir, Konya, Kütahya ve Tekirdağ olmak üzere 21 farklı şehirde görev yapan öğretmen araştırmaya katılmıştır. Benzer şekilde öğretmenlerin branşları, çalışma süreleri, çalıştıkları okul düzeyleri de değişkenlik göstermektedir. Katılımcıların %24.3'ü matematik, %20'si fen bilimleri, %18.5'i sınıf öğretmenliği, %12.9'u İngilizce, %11.4'ü Türkçe, %10'u sosyal bilgiler ve %2.9'u din bilgisi branş öğretmenleridir. Okul düzeyi göz önünde bulundurulduğunda ise öğretmenlerin %22.8'i ilkokulda, %62.9'u ortaokulda, ve %14.3'ü lisede görev yapmaktadırlar. Çalışma süreleri 2 yıl ile 31 yıl arasında değişmekte olup ortalama görev süreleri 10.4 yıldır. Araştırmaya katılan öğretmenlerin %97.2'si devlet okullarında görev yaparken %2.8'i özel okulda görev yapmaktadır.

Ölçme Aracı ve Verilerin Analizi

Araştırmada öğretmenlerin Sato Test Kuramı ile yapılan sınıflamalara yönelik görüşlerinin alınması için anket kullanılmıştır. Bu çalışmanın araştırmacıları tarafından hazırlanan anket maddeleri öncelikle dil ve anlatım yönünden problem teşkil edecek herhangi bir durum olup olmadığını belirlemek amacıyla 5 öğretmenin görüşlerine sunulmuş, alınan dönütler doğrultusunda maddelere son halleri verilmiştir. Anket maddelerinin yanıtlanmasına geçmeden önce katılımcılara, 10 maddelik bir teste ait 20 kişilik öğrenci grubunun Sato Kuramına göre analiz edilmiş verileri sunulmuştur. Bu bilgi ile Sato Kuramının madde ve öğrenciler için yaptığı sınıflamalar ve bu sınıflamalara karşılık gelen belirlemeler (dönütler) öğretmenlere açıklanmış ve bu kuram ile alışık oldukları Klasik Test Kuramı arasındaki benzerlik ve farklılıklara dikkat çekilmiştir. Bu bilgilendirmeden sonra, öğretmenlerden konuya ilişkin görüşlerini yapılandırılmış anket maddeleri üzerinden yansıtmaları istenilmiştir. Söz konusu ankette açık uçlu sorular dışındaki sorular için *Evet*, *Kararsızım*, *Hayır* şeklinde üç kategori kullanılmıştır. Çalışmada kullanılan anket 12 maddeden oluşmaktadır. 70 katılımcıdan elde edilen verilere frekans analizi uygulanmış, her bir madde için elde edilen frekans ve yüzde değerleri ayrı ayrı raporlanmıştır. Bu çalışmada, maddelerin işlevliliği ve öğrencilerin öğrenme durumları hakkında sınıflama yapmak için kullanılan madde ve öğrenci uyarı indeksleri Sheu, Pham, Nguyen ve Nguyen (2013) tarafından Öğrenci-Madde Çizelgesi Analizi ve Gri Öğrenci-Madde Çizelgesi Analizi için geliştirilen Matlab paketi ile elde edilmiştir. Paket, Matlab R2013b versiyonu üzerinde çalıştırılmıştır.

Bulgular ve Yorum

Araştırma kapsamında kullanılan 12 anket maddesine 70 katılımcının verdiği yanıtlar, frekanslar ve yüzde oranlar aracılığıyla incelenmiştir. Elde edilen bulgular Tablo 1’de sunulmuştur.

Tablo 1

Ankette Yer Alan Maddelere Verilen Yanıtların Frekans ve Yüzde Dağılımları

Madde No	Maddeler	Evet		Kararsızım		Hayır	
		N	%	N	%	N	%
1	Uygulayacağım testlerden elde ettiğim sonuçları Sato Test Kuramı’na dayalı olarak analiz etmek isterim.	57	81.4	2	2.9	11	15.7
2	Testlerde kullanacağım maddeler ile ilgili sözel (nitel) dönütler almak işime yarar.	67	95.7	1	1.4	2	2.9
3	Test maddeleri ile ilgili verilen dönütleri beğendim.	58	82.9	10	14.3	2	2.8
4	Test maddeleri ile ilgili verilen dönütleri yeterli buldum.	49	70	18	25.7	3	4.3
5	Test maddeleri ile ilgili sözel dönütler almam madde havuzu oluşturma konusunda işimi kolaylaştırır.	62	88.6	8	11.4	-	-
6	Test maddeleri ile ilgili sözel dönütler almak madde havuzu (soru havuzu) oluşturmam için beni teşvik eder.	61	87.1	7	10	2	2.9
7	Klasik yöntemlerle madde analizi yapmayı tercih ederim.	2	2.9	29	41.4	39	55.7
8	Öğrencilerle ilgili sözel (nitel) dönütler almak işime yarar.	62	88.6	5	7.1	3	4.3
9	Öğrencilerle ilgili verilen dönütleri beğendim.	56	80	12	17.1	2	2.9
10	Öğrencilerle ilgili verilen dönütleri yeterli buldum.	45	64.3	22	31.4	3	4.3
11	Öğrencilerle ilgili alacağım dönütler öğretimi iyileştirme (düzenleme) konusunda işimi kolaylaştırır.	62	88.6	7	10	1	1.4
12	Öğrencilerin test puanları benim için yeterli bir dönüttür ve sözel dönütler gereksizdir.	7	10	7	10	56	80

Tablo 1’de, arařtırmaya katılan öğretmenlerin anket maddelerine verdikleri yanıtların frekans ve yüzde dağılımları görölmektedir. İlk maddeye verilen yanıtlar incelendiğinde; katılımcıların “evet” yanıtı verenlerin %81.4, “kararsızım” yanıtını verenlerin %2.9 ve “hayır” yanıtı verenlerin oranının ise %15.9 olduđu görölmektedir. Elde edilen bulgular, katılımcıların büyük çoğunluğunun uygulayacakları testlerde STK’dan faydalanmak istediğini göstermektedir.

İkinci maddeye verilen yanıtlar incelendiğinde; katılımcıların %95.7’sinin “evet”, %1.4’ünün kararsızım ve %2.9’unun ise “kararsızım” yanıtını verdikleri görölmektedir. Katılımcıların neredeyse tamamının bu maddeye “evet” yanıtı vermesinin dikkat çekici olduđu söylenebilir. Bu durum, öğretmenlerin, maddelerle ilgili verilen dönütlerin işlerine yarayacağını düşündükleri şeklinde yorumlanabilir.

Üçüncü maddeye verilen yanıtlar incelendiğinde; katılımcıların %82.9’unun test maddeleri ile ilgili verilen dönütleri beğendiği, %2.8’inin beğenmediği görölmektedir. Bu konuda kararsız olanların oranı ise %14.3’tür.

Dördüncü maddeye verilen yanıtlara bakıldığında; katılımcıların %70’inin test maddeleri ile ilgili verilen dönütleri yeterli buldukları, %4.3’ünün ise bu dönütlerin yetersiz olduğunu düşündükleri görölmektedir. Bu konuda kararsız olanların oranı %25.7 olarak elde edilmiştir. Bu bulgu, bir önceki madde ile birlikte değerlendirildiğinde, katılımcıların bir kısmının dönütleri beğenmekle birlikte bu dönütlerin yeterliliği ile ilgili kararsız oldukları yorumu yapılabilir.

Beşinci maddeye verilen yanıtlar incelendiğinde; katılımcıların %88.6’sının test maddeleri ile ilgili verilen dönütleri madde havuzu oluşturma noktasında işe yarar bulduđu, kalan %11.4’lük kesimin ise bu konuda kararsız olduđu görölmektedir. Bu bağlamda,

dönütlerin madde havuzu oluşturma noktasında işe yaramayacağı yönünde görüş belirten hiçbir katılımcının olmadığı ayrıca vurgulanabilir.

Altıncı maddeye verilen yanıtlar incelendiğinde; katılımcıların %87.1'inin alacakları dönütlerin madde havuzu oluşturma sürecinde teşvik edici olduğunu düşündükleri görülmektedir. Bu konuda kararsız olanların oranı %10 iken, “hayır” diyenler %2.9'luk bir kesimi oluşturmaktadır.

Anketin yedinci maddesi olan “Klasik yöntemlerle madde analizi yapmayı tercih ederim.” ifadesine verilen yanıtlar incelendiğinde; “evet” yanıtı verenlerin oranının %2.9, “kararsızım” diyenlerin oranının %41.4 ve “hayır diyenlerin oranının ise %55.7 olduğu bulgusuna ulaşılmıştır. Bu madde, tüm anket maddeleri içerisinde “kararsızım” yanıtı verenlerin en yüksek orana sahip olduğu madde olarak karşımıza çıkmaktadır. Bu bulgu, katılımcıların büyük bir kısmının, maddelerle ilgili verilen dönütleri önemli ve işe yarar bulmakla birlikte STK'yı klasik yöntemlere tercih etmek yönünde kararsız oldukları şeklinde de yorumlanabilir. Ayrıca, bu maddeye “evet” yanıtı veren iki katılımcıya klasik yöntemleri tercih etmelerinin nedenleri sorulmuş ancak bu katılımcılar yanıtlarına ilişkin herhangi bir görüş belirtmeye gerek görmemiştir.

Sekizinci maddeye verilen yanıtlar incelendiğinde; katılımcıların %88.6'sının öğrencilerle ilgili nitel dönütler almanın işlerine yarayacağını düşündükleri görülmektedir. Bu konuda kararsız olanların oranı %7.1 iken, kalan %4.3'lük kesim öğrencilerle ilgili dönütlerin işlerine yaramayacağı görüşündedirler.

Dokuzuncu madde için elde edilen yanıtlara bakıldığında; katılımcıların %80'inin öğrencilerle ilgili verilen dönütleri beğendiği görülürken, %17.1'inin bu konuda kararsız olduğu, kalan %2.9'unun ise “hayır” yanıtı verdiği görülmektedir.

Araştırma kapsamında kullanılan anket maddelerinin onuncusuna verilen yanıtlar incelendiğinde; “Öğrencilerle ilgili verilen dönütleri yeterli buldum.” ifadesine “evet” yanıtı verenlerin oranının %64.3, “hayır” diyenlerin oranının ise %4.3 olduğu görülmektedir. “Test maddeleri ile ilgili verilen dönütleri yeterli buldum.” ifadesine (4.madde) kararsız olduğu yönünde tepki verenlerin oranına benzer olarak; bu maddeye “kararsızım” yanıtı verenlerin oranının da (%31.4) diğer maddelerdeki kararsızlara nazaran yüksek olduğu görülmektedir.

On birinci maddeye verilen yanıtlar incelendiğinde, katılımcıların %88.6’sının STK’yı kullanmaları durumunda öğrencilerle ilgili alacakları dönütlerin öğretimi iyileştirme noktasında işe yarayacağını düşündükleri belirlenmiştir. Bu maddeye “kararsızım” yanıtı verenlerin oranının %10, “hayır” yanıtı verenlerin oranının ise %1.4 olduğu görülmektedir.

Anketteki son madde olan “Öğrencilerin test puanları benim için yeterli bir dönüttür ve sözel dönütler gereksizdir.” maddesine katılımcıların %80’i “hayır” yanıtını vermişlerdir. Kalan %20’lik kesimin yarısı “evet” yanıtını vermiş, diğer yarısı ise “kararsızım” demiştir.

Anketi tamamlamalarının ardından katılımcılara, konuya ilişkin paylaşmak istedikleri başka bir görüşlerinin olup olmadığı sorulmuştur. Katılımcılardan dördü bu kısımda görüşlerini paylaşmıştır. Bu katılımcıların, önemli bir konuda çalışma yürütüldüğü ve bu tür çalışmaların eğitimde ölçme ve değerlendirme süreçlerinde yaşanan sorunların aşılabilmesi bakımından gerekli olduğu görüşünde birleştikleri görülmüştür.

Sonuç ve Tartışma

Bu çalışmada, Sato Test Kuramı ile yapılan sınıflandırmalara yönelik öğretmen görüşlerinin toplanması ve incelenmesi amaçlanmıştır. Bu amaçla araştırmacılar tarafından oluşturulan anket 21 farklı şehirde görev yapan 70 öğretmene uygulanmış, anket maddelerine verilen yanıtlar frekans ve yüzde oranlar aracılığıyla incelenmiştir. Elde edilen bulgular,

katılımcı öğretmenlerin büyük çoğunluğunun (%81.4) STK'yı kullanmaya istekli olduklarını göstermektedir. Bununla birlikte katılımcıların, STK ile test maddelerinin işlevliliği ve öğrencilerin öğrenme durumları hakkında verilen dönütleri beğendikleri ve işe yarar buldukları sonucuna ulaşılmıştır. Bu bakımdan, araştırma sonuçlarının Wang ve Chen'in (2013) Tayvanlı öğretmen ve öğrencilerin STK dönütlerine yönelik görüşlerini inceledikleri araştırmanın sonuçlarıyla benzerlik gösterdiği ifade edilebilir. Buna karşın, bu araştırmaya katılan öğretmenlerin önemli bir kısmının maddeler ve öğrencilerle ilgili dönütlerin yeterli olup olmadığına ilişkin kararsızlık yaşadıkları görülmüştür. Ayrıca katılımcıların %41.4'ünün madde analizleri için klasik yöntemleri tercih etme noktasında kararsızlık yaşadıklarını ifade etmeleri de dikkat çekici bir sonuç olarak nitelendirilebilir. Bu durumun, öğretmenlerin büyük çoğunluğunun dönütleri beğendiği, madde havuzu oluşturma konusunda teşvik edici bulunduğu ve öğretimi iyileştirme noktasında işlerini kolaylaştıracağını düşündüğü yönünde elde edilen bulgularla birlikte yorumlanmasının doğru olacağı düşünülmektedir. Aynı öğretmenlerin STK'yı tercih etme noktasında yaşadıkları kararsızlığın daha önce eğitimini almadıkları ve ölçme süreçlerinde kullanmadıkları bir kuramla ilk defa karşılaşmalarından kaynaklı olarak ortaya çıkabileceği iddiası öne sürülebilir. Bu anlamda, konuya ilgi duyan araştırmacılara öğretmenlerin belirlenen bir süre boyunca STK'yı sınıf içi ölçme ve değerlendirme süreçlerinde deneyimleyebilecekleri bir ortam sunmaları ve buna uygun bir araştırma düzeyi tasarımları yönünde bir öneri getirilebilir. Yapılacak olan araştırmalarda kullanılacak testler kağıt-kalem formatında olabileceği gibi ölçme süreçlerinde web-tabanlı sistemlerden de faydalanılabilir.

Kaynakça

- Acar, T. (2006). *Sato uyarı indeksleri ile madde ve başarı analizleri*. [Çevrimiçi: [http://www.academia.edu/11390990/Sato Uyarı İndeksleri ile Madde ve Başarı A nalizleri](http://www.academia.edu/11390990/Sato_Uyarı_İndeksleri_ile_Madde_ve_Başarı_A_nalizleri), Erişim tarihi: 4 Mayıs 2017.]
- Bayrak, F. ve Yurdugül, H. (2016). Web-tabanlı öz-değerlendirme sisteminde öğrenci uyarı indeksini temel alan öğrenme analitiği modülünün tasarlanması. *Eğitim Teknolojisi Kuram ve Uygulama*, 6(2), 85-99.
- Birnbaum, A. (1957). *Efficient design and use of tests of a mental ability for various decision-making problems*. Series Report No. 58-16. Project No. 7755-23, USAF School of Aviation Medicine, Randolph Air Force Base, Texas: January.
- Birnbaum, A. (1958). *Further considerations of efficiency in tests of a mental ability*. Technical Report No. 17. Project No. 7755-23, USAF School of Aviation Medicine, Randolph Air Force Base, Texas.
- Crocker, L. ve Algina J. (1986). *Introduction to classical and modern test theory*. Orlando: Harcourt Brace Jovanovich Inc.
- Çelen, Ü. ve Aybek, E.C. (2013). Öğrenci başarısının öğretmen yapımı bir testle klasik test kuramı ve madde tepki kuramı yöntemleriyle elde edilen puanlara göre karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 4(2), 64-75.
- Çüm, S., Gelbal, S. ve Tsai C.P. (2016). Sato test kuramı yöntemleriyle farklı örneklemelerden elde edilen madde parametrelerinin tutarlılığının incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 7(1), 170-181.

de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.

DeMars, C. (2016). *Madde tepki kuramı*. H. Keleciođlu (Çev. Ed.). Ankara: Nobel.

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

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

Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/response statistics. *Educational and Psychological Measurement*, 58(3), 357-381.

Hambleton, R. K. ve Swaminathan, H. (1985). *Item response theory: Principles and applications*, New York: Springer.

Hambleton, R. K., Swaminathan, H., ve Rogers, H. (1991). *Fundamentals of item response theory*. CA: Sage.

Hulin, C. L., Lissak, R. I., ve Drasgow, F. (1982). Recovery of two and three-parameter logistic item characteristic curves: A Monte Carlo study. *Applied Psychological Measurement*, 6(3), 249-260.

Kiany, G.R. ve Jalali, S. (2009). Theoretical and practical comparison of classical test theory and item response theory. *International Journal of Active Learning*, 12(1), 167-197.

- Lin, Y.H., ve Yih, J.M. (2015). Application of IIRS in mathematics instruction to promote pupils decimal concept. *The International Conference on Language, Education and Psychology*, Taiwan.
- Lord, F. ve Novick R.M. (1968). *Statistical theories of mental test scores*. New York: Addison Wesley Publishing Company.
- MacDonald, P. ve Paunonen, S. (2002). A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and Psychological Measurement*, 62(6), 921-943.
- McArthur, D.L. (1983). *Analysis of test score patterns: The student-problem (s-p) technique (Report no: 218)*. Los Angeles: Center for the Study of Evaluation, University of California.
- Pham, D.H., Sheu, T.W., ve Nagai, M. (2015). PCSP 1.0 software for partial credit S-P chart analysis. *International Journal of Hybrid Information Technology*, 8(6), 309-322.
- Ree, M. J., ve Jensen, H. E. (1983). *Effects of sample size on linear equating of item characteristic curve parameters: Latent trait test theory and computerized adaptive testing*. New York: Academic Press.
- Sato, T. (1984). *The state of art on S-P analysis activities in Japan*. Tokyo: NEC Corp.
- Sheu, T.W., Pham, D.H., Nguyen, P.T., ve Nguyen, P.H. (2013). Amatlab toolbox for student-problem chart and grey student-problem chart and its application. *International Journal of Kansei*, 4(2), 75-86.

- Sheu, T.W., Pham, D.H., Tsai, C.P., Nguyen, P.T., Nguyen, P. H. ve Nagai, M. (2014a). Rasch GSP toolbox for assessing academic achievement. *Journal of Software*, 9(7), 1903-1913.
- Sheu, T.W., Nguyen, P.T., Pham, D.H., Tsai, C.P. ve Nagai, M. (2014b). A MATLAB toolbox for misconceptions analysis based on S-P chart grey relational analysis and ROC. *Transactions on Machine Learning and Artificial Intelligence*, 2, 72-85.
- Suen, H.K. (1990). *Principles of test theories*. Hillsdale, NJ: [Lawrence Erlbaum Associates](#).
- Switzer, D. M. ve Connell, M. L. (1990). Practical applications of student response analysis. *Educational Measurement: Issue and Practice*, 9(2), 15-17.
- Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurements Yearbook*, 54, 837–847.
- Traub, R. ve Rowley, G.(1991). Understanding reliability. *Educational Measurement: Issues and Practice*, 8(1), 8-14.
- Tsai, C.P., Sheu, T.W., Tzeng, J.W.,Chen, H.J., Chiang, H.J. ve Nagai, M. (2014). Diagnose learning misconceptions based on rough sets. *International Journal of Applied Mathematics and Statistics*, 52(2), 63-75.
- van der Linden, W. J. ve Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer.
- Wang, C.H. ve Chen, C.P. (2013). Employing online S-P diagnostic table for qualitative comments on test results. *The Electronic Journal of e-Learning*, 11(3), 263-271.
- Yu, M. N. (2011). *Educational testing and assessment*. Taiwan: Psychology Publisher.

Extended Abstract

Introduction

Psychometricians mainly deal with the following issues on the basis of mathematical and statistical theories, which can be called “test theory” or “psychometric theory”: development and selection of test items; the methods used to express responses / replies in numerical terms; and the interpretation and evaluation of the obtained results. Among these test theories, Classical Test Theory (CTT) and Item Response Theory (IRT) have been studied the most. The related literature involves a large number of studies discussing the strengths and weaknesses of CTT and IRT. Sato Test Theory (STT), which was suggested by Takahiro Sato in 1969, was built on the analysis methods which are expected to produce accurate results in determining learning success of students by examining their response patterns even in small-size samplings which do not display normal distribution. Although STT was originally used to classify students according to their learning success and categorize the test items according to their quality, it was later used to determine achievement / ability levels of students thanks to Grey Relationship Analysis developed by Ju-Long Deng based on Nagai’s suggestion.

Aim

STT is an important measurement and evaluation method used in Japanese primary and secondary schools (Sato, 1984). However, it isn’t popular enough in Turkey and very few studies focusing on this theory have been conducted so. Since STT claims to obtain valid and reliable results in educational environments with frequent measurement and evaluation practices, it is important to explore the opinions of not only field experts or researchers as test developers but also regular school teachers regarding the theory. Therefore, this study aims to obtain data regarding the opinions of teachers about the classifications made by using STT

and to examine the data obtained. It is believed that such studies will lead to an increasing interest in the theory, and further studies will be conducted accordingly.

Study Group

The study group consists of 70 volunteer teachers. The survey developed by the researchers was administered to 51 participants online and 19 participants face-to-face.

Method

The study uses a descriptive method to collect the data about the opinions of teachers regarding the classifications made by using STT. For the purposes of data collection, the researchers developed a 12-item survey. Prior to the administration of the survey, the participants were presented the data collected from 20 students through a 10-item test and analyzed according to STT. The aim of this presentation was to provide information about the classifications made by using STT for both students and items. It also presented the feedback about these classifications and highlighted the similarities and differences between this theory and CTT. After this introduction, the teachers were asked to reflect their opinions about the issue by replying the items in the survey. The data obtained were examined by calculating frequencies and percentages.

Findings

The results of the study show that majority of participant teachers (81.4%) are eager to use STT. In addition, it was found that they liked the feedback given by STT about the effectiveness of test items and the learning success of students. Moreover, most of the participant teachers found such feedback encouraging in terms of item pool preparation and thought that it might play important role in improving learning. However, a considerable number of participants reported that they found it difficult to decide whether this feedback is

sufficient or not. Finally, 41.4 % of the participants were indecisive about using classical methods for item analysis or not, which is an outstanding finding.

Conclusion

Since teachers displayed a positive attitude towards STT feedback, we can conclude that the results of the study are consistent with those of the study by Wang and Chen (2013), which examined the opinions of Taiwanese teachers and students about STT feedback. In addition, the majority of the participants found such feedback about items important and useful but they were indecisive about preferring STT to classical methods. This finding may imply that teachers may hesitate using a method they have never heard or trained about for measurement and evaluation purposes. Thus, further studies might be designed in a way that researchers provide participant teachers with an environment in which they can experience STT in terms of in-class measurement and evaluation processes.