



Sınıf İçi Ölçme ve Değerlendirmede Puanlara Anlam Kazandırma: %70 Doğru Yanıt Yöntemi

Proficiency-Based Classroom Assessment: 70% Correct Response Method

Serkan Arıkan, Muğla Sıtkı Koçman Üniversitesi, serkanarikan@mu.edu.tr
Sevilay Kilmen, Abant İzzet Baysal Üniversitesi, sevilaykilmen@ibu.edu.tr

Öz. Bu çalışmanın amacı sınıf içi ölçme ve değerlendirme uygulamaları kapsamında geliştirilen testler temel alınarak, %70 doğru yanıt yöntemi ile puanlara anlam kazandırılıp kazandırılmayacağını araştırmaktır. Bu amaç doğrultusunda %70 doğru yanıt yöntemiyle elde edilen soru gruplarının bilişsel taksonomiye göre sıralanıp sıralanmadığı ve %70 doğru yanıt yöntemi ile elde edilen soru grupları ile madde haritalama yöntemi sonucunda elde edilen soru gruplarının uyumu incelenmiştir. Araştırmanın bulguları %70 doğru yanıt yöntemi kullanılarak elde edilen soru gruplarının, testin geliştirilmesinde temel alınan TIMSS matematiksel düşünme becerileri gibi hiyerarşik bir yapı gösterdiğini ortaya koymaktadır. Ayrıca bulgular %70 doğru yanıt yöntemi ve madde haritalama yönteminin belli ölçüde benzerlik gösterdiğine işaret etmektedir. Bu yöntem sayesinde öğretmenlerin öğrencilere ve velilere test sonuçlarını kullanarak öğrencilerin bilgi ve becerileri hakkında daha sık anlamlı geri bildirim verebilecekleri düşünülmektedir.

Anahtar Sözcükler: %70 doğru yanıt yöntemi, puanlara anlam kazandırma, madde haritalama yöntemi, bilişsel süreçler

Abstract. The aim of this study was to investigate whether the classroom test scores could be used to give proficiency-based results by using the 70% correct response method. For this purpose, the extent to which the question groups obtained with 70% correct response method represented the proposed cognitive dimensions and the similarity of the question groups obtained with 70% correct response method and item mapping method were examined. The findings of the study showed that the question groups obtained using the 70% correct response method were hierarchical. In addition, findings indicated that the 70% correct response method and item mapping method showed some similarities. With the use of this method, it is believed that teachers could give more frequent proficiency-based feedback on their test results.

Keywords: 70% correct response method, Proficiency-based feedback, Item mapping, Cognitive dimensions

SUMMARY

Introduction

Defining student proficiencies is an important factor in increasing the quality of teaching by making accurate and concrete planning about the knowledge and skills to be developed. Having only exam grades is not enough to make the suitable teaching plan. At this point, proficiency definitions are needed. Ability differences between the students who gets different scores can be demonstrated by assigning meanings to these scores. A teacher who has information about the proficiency of his or her classroom can better plan the lesson and perform the teaching more effectively.

The aim of this study was to investigate whether it is possible to give proficiency-based results using the 70% correct response method, described by Berberoglu (2006), based on the classroom tests. For this purpose, the extent to which the question groups obtained with 70% correct response method represented the proposed cognitive dimensions and the similarity of the question groups obtained with 70% correct response method and item mapping method were examined.

Method

Since this study aimed to assign meaning to the scores obtained from classroom tests via proficiency descriptions, a classroom test that was developed by the last grade preservice mathematics teachers within the test development courses was used. The test was administered to the 8th grade middle school students. Fourty-three senior mathematics teacher candidates, by creating 8 teams, developed classroom tests for 8th grade students. Each test was applied to 100 different 8th grade students. From these tests, a test with satisfactory item parameters was chosen with the aim of exemplifying the 70% correct response method. The selected test in this study was developed by a team of 6 teacher candidates. The process of assigning meaning to the scores was described here. Firstly, student scores were grouped into 5 categories (1-20; 21-40; 41-60; 61-80; 81-100). The number of students in each group was counted and 70% of these numbers in each group was calculated. Then, for each score category, questions that were answered correctly by 70% or more of students in these groups were listed. Same procedure was replicated for the other score categories as well. Afterwards, proficiency definitions were decided based on verbal definitions describing the skills measured by the questions. ConstructMap 4.6 (Kennedy, Wilson, Draney, Tutuncuyan, & Vorp, 2010) program was used to determine the question groups according to the item mapping method, showing the total raw score of the students, the student ability and the item difficulty values in the Wright map.

Results

The findings of the study showed that the question groups obtained using the 70% correct response method were hierarchical. When examining the proficiency definitions obtained using question groups, a hierarchical structure similar to knowing, applying and reasoning dimensions of TIMSS mathematical thinking skills was emerged. In addition, findings indicated that the 70% correct response method and item mapping method showed some similarities. Item mapping with Wright map is based on a 50% probability of correctly responding to questions. For this reason, a question at a certain level in the Wright map method was expected to be at the same or higher level in the 70% correct response method. While nine questions were at exactly the same level, nine questions were at a higher level. Although there were differences in the calculation methods, it was thought that there was a similarity between the question groups obtained by both methods.

Discussion and Conclusion

There is a need for a method that will enable teachers to define proficiency level of students using classroom test results. In this study, the 70% correct response method was evaluated for applicability with real student data, and similarity of the results with item response theory item mapping.

These results showed that teacher candidates could assign meaning to the test scores with an accurate orientation. This method is certainly not an alternative to the proficiency definitions set by experienced experts using item response theory. However, for teachers who cannot get professional help from measurement experts, this method might provide valuable information in addition to score report. This study showed that 70% correct response method is justifiable and usable.

GİRİŞ

Ölçme ve değerlendirmenin en önemli amaçlarından biri anlamlı geri bildirim sağlamasıdır (Berberoğlu, 2006; Berberoğlu, 2009; Demirtaşlı, 2017). Anlamlı geri bildirimden öğrencilere yeterlik tanımları (sözel bilgi ve beceri tanımları) yardımı ile hangi bilgi ve becerilere sahip olduklarının iletilmesi ve varsa öğrencilerin eksik oldukları bilgi ve becerilerin belirlenmesi anlaşılmaktadır (Arıkan, 2017). Bir sınavdan 45 alan öğrenci ile 85 alan öğrenci arasında bilgi ve beceri bakımından ne gibi farklar bulunmaktadır? Eğer öğrenci yeterlikleri bilinirse, öğrencilerin sahip oldukları bilgi ve becerilerden başlayarak, geliştirilmesi gereken bilgi ve beceriler hakkında doğru ve somut bir planlama yapılarak öğretimin kalitesini artırmak mümkün olabilir. Sadece sınav notlarına sahip olmak doğru iyileştirme planını yapabilmek için yeterli olamayacaktır. Yeterlik tanımları sayesinde belirli puan aralığında yer alan öğrencilerin sahip oldukları ve geliştirmeleri gereken bilgi ve beceriler belli olacağı gibi, farklı puan aralığında yer alan öğrenciler arasındaki farklar da ortaya konmuş olacaktır. Sınıfının yeterlikleri hakkında bilgi sahibi olan bir öğretmen dersini daha iyi planlayabilecektir. Daha etkili öğretim için öğretmenlerin öğrenci yeterliklerini elde edilebildiği ölçme ve değerlendirme uygulama ve sonuçlarına ihtiyaçları vardır.

Öğrenci puanlarına karşılık gelen yeterliklerin tanımlandığı PISA (The Programme for International Student Assessment; Uluslararası Öğrenci Değerlendirme Programı) ve TIMSS (Trends in International Mathematics and Science Study; Uluslararası Matematik ve Fen Eğilimleri Araştırması) gibi uluslararası geniş ölçekli uygulamalar mevcuttur (Mullis, Martin, Foy, & Hooper, 2016; OECD, 2016). PISA ve TIMSS amacı gereği öğrenci bazında değil ülkeler bazında ilgili yeterlik düzeylerinde yer alan öğrenci dağılımlarını raporlamaktadır. Öğrenci bazında bu yeterlikler kullanılmamaktadır (Rutkowski, Gonzalez, Joncas, von Davier, 2010). Türkiye’de öğrencilerin yeterlik düzeyinin belirlenmesine yönelik çalışmalara ise Cito Türkiye Öğrenci İzleme Sistemi örnek verilebilir. Cito Türkiye Öğrenci İzleme Sistemi’nde dijital sınavlar yapılarak öğrenci yeterlikleri sözel tanımlayıcılar yardımı ile raporlaştırılmıştır (Özgen Tuncer, 2009). Milli Eğitim Bakanlığı tarafından, ABİDE (Akademik Becerilerin İzlenmesi ve Değerlendirilmesi) uygulaması Türkiye’de gerçekleştirilen öğrenci yeterliklerini belirleme çalışmalarına verilebilecek diğer bir örnektir (MEB, 2017). Bu uygulamalar modern test teorisini temel alarak büyük gruplar için yeterlik tanımlarını belirleme açısından yol gösterici çalışmalardır. Bu çalışmalar yeterlik tanımlarını temel alarak mutlak değerlendirme yapmanın önemini ve kullanılabilirliğini örneklemiştir.

Yeterlik Tanımlarını Belirleme

Bu bölümde PISA, TIMSS ve ABİDE kapsamında yeterliklerin nasıl belirlendiği özetlenmektedir.

PISA Kapsamında Yeterlik Tanımlarının Belirlenmesi

PISA, 2015 uygulamasından itibaren madde kestirimleri için çoktan seçmeli veya 0-1 şeklinde puanlanan sorular için iki parametrelili lojistik modeli, ikiden fazla puan kategorisi olan sorularda ise genelleştirilmiş kısmi puan modelini (generalized partial credit model) kullanmaya başlamıştır. Elde edilen madde güçlük parametreleri temel alınarak madde haritaları oluşturulmakta ve uzmanlar tarafından 6 kesme noktası ve 7 yeterlik düzeyi belirlenmektedir. Madde tepki kuramının madde güçlüğü tanımıyla ilişkili olarak, bir yeterlik düzeyinin en altında yer alan soruların o düzeyin en altında bulunan öğrenciler tarafından %50 olasılıkla doğru yanıtlanması beklenirken, aynı yeterlik düzeyinde olan ama daha üst sıralarda yer alan öğrencilerce bu soruların daha yüksek olasılıkla doğru yanıtlanması beklenmektedir (OECD, 2017).

TIMSS Kapsamında Yeterlik Tanımlarının Belirlenmesi

TIMSS 2015 kapsamında madde parametreleri kestirilirken çoktan seçmeli maddeler için 3 parametrelili lojistik model, 0-1 şeklinde puanlanan açık uçlu sorular için 2 parametrelili lojistik model ve 0-1-2 gibi 2 kategoriden fazla puanlamanın yapıldığı açık uçlu sorularda ise kısmi puan modeli (partial credit model) kullanılmıştır. TIMSS’de 4 kesme puanı kullanılarak 5 yeterlik

düzeyi (alt düzey altı, alt düzey, orta, üst ve ileri düzey) belirlenmektedir. Bu kesme puanları 400, 475, 550 ve 625 olarak önceden belirlenmiştir. Ardından çoktan seçmeli sorular için %65, açık uçlu sorular için %50 yanıtlanma yüzdesine göre sorular her düzeyle eşleştirilmektedir. Örneğin, çoktan seçmeli bir soru “alt düzeyde” yer alan öğrenciler tarafından %65 veya üstü yüzde ile doğru yanıtlanıyor ise bu düzeydeki tanımlamalara alınmaktadır. Diğer bir soru “orta düzeyde” yer alan öğrenciler tarafından %65 veya üstü yüzde ile doğru yanıtlanıyor ama “alt düzeyde” %50 veya altı yüzde ile doğru yanıtlanıyor ise bu soru “orta düzeydeki” yeterlik tanımlarına alınmaktadır. Bu şekilde devam ederek, tüm sorular ilgili yeterlik düzeyleri ile ilişkilendirilmektedir. Ardından uzmanlar tarafından soruların ölçtükleri bilgi ve beceriler yeterlik tanımı haline getirilmektedir (Martin, Mullis & Hooper, 2016).

ABİDE Kapsamında Yeterlik Tanımlarının Belirlenmesi

ABİDE kapsamında işaretleme (bookmark) yöntemi kullanılarak yeterlik düzeyleri belirlenmiştir. Madde parametre kestirimleri çoktan seçmeli maddeler için 3 parametrelili lojistik model, 0-1 şeklinde puanlanan açık uçlu sorular için 2 parametrelili lojistik model ve 0-1-2 gibi 2 kategoriden fazla puanlamanın yapıldığı açık uçlu sorularda ise aşamalı tepki modeli (graded response model) kullanılarak yapılmıştır. İşaretleme işlemi için her ders alanında 15 öğretmen (5 grup) görev almıştır. Görev alan öğretmenlere sınav sorularının kolaydan zora doğru sıralanarak yeniden oluşturulmuş kitapçıklar ile madde haritaları dağıtılmıştır. Gruplar halinde çalışan öğretmenler belirli tartışmalar sonucunda dört işaretleme (kesme noktası) yapmışlardır. Her grubun yaptıkları işaretlemelerin ortancaları alınarak nihai kesme noktaları belirlenmiştir. Ardından her düzeye denk gelen sorular incelenerek yeterlik tanımları düzenlenmiştir. Yeterlik düzeyleri, temel altı, temel, orta, orta üstü ve ileri olarak adlandırılmıştır (MEB, 2017).

Çalışma Hakkında

Geniş ölçekli uygulamalar yapılması gereken ideal ölçme ve değerlendirme yaklaşımlarını sınıflardaki öğretmenlere gösterse de öğretmenlerin sınıflarında kullandıkları ölçme uygulamaları bu yaklaşımlardan oldukça uzaktır. Öğretmenler sınavları büyük oranda not verme amacı ile kullanmakta, öğrencilerine geri bildirim nadiren vermekte, yeterlik tanımlarını ise, sahip olmadıkları veya oluşturamadıkları için, kullanmamaktadırlar. Türkiye’deki öğretmenler ise ölçme ve değerlendirme sonuçlarını öğrencilerine geri bildirimde bulunmak amacı ile Avrupa Birliği üyesi ülkelere oranla çok az sıklıkta kullanmaktadır (Demirtaşlı, 2009). PISA, TIMSS ve ABİDE gibi uygulamaların ortak özelliği geniş ölçekli uygulamalar olmalarının yanı sıra psikometri ve alan uzmanlarından oluşan ekipler ile ileri istatistiksel yöntemleri kullanarak yeterlik tanımlarını oluşturmalarıdır. Öğretmenlerin uzman bir ekip desteği alamayışı ve bu türden istatistikler teknikler konusunda bilgi sahibi olmayışı yeterlik belirleme çalışmalarını engellemektedir. Buradan hareketle öğretmenlerin sınıf içi uygulamalarda kullanabilecekleri pratik yeterlik belirleme yöntemine ihtiyaç duyulmaktadır. Bu ihtiyaca yönelik olarak, Berberoğlu (2006) puanlara nasıl anlam kazandırılabilirliğini göstermek amacı ile kitabının puan analizi bölümünde bir yöntemin teorik yapısını aktarmıştır. Bu çalışmada bu yöntem “%70 doğru yanıt yöntemi” olarak adlandırılmaktadır. %70 doğru yanıt yöntemi öğretmenlerin öğrencilerin doğru yanıtlama yüzdelerini inceleyerek soruları gruplayabilmelerini sağlayan bir yöntemdir. Detayları yöntem kısmında aktarılmaktadır.

Bu çalışmanın amacı sınıf içi ölçme ve değerlendirme uygulamaları kapsamında geliştirilen testler ve %70 doğru yanıt yöntemi kullanılarak yeterlik tanımları bazında puanlara anlam kazandırılıp kazandırılmayacağını araştırmaktır. Bu amaçla, çalışmada alttaki araştırma soruları yanıtlanmaktadır.

1) %70 doğru yanıt yöntemiyle elde edilen soru grupları bilme, uygulama ve akıl yürütme bilişsel basamaklarına göre sıralanmakta mıdır?

2) Öğretmenlerin %70 doğru yanıt yöntemi ile elde ettikleri soru grupları ile madde tepki kuramı sonucu elde edilen soru grupları uyumlu mudur?

YÖNTEM

Çalışma Grubu

Bu çalışma sınıf içi ölçme ve değerlendirme araçlarınca elde edilen puanlara yeterlik tanımlarını temel alarak anlam kazandırmayı amaçladığı için, son sınıf matematik öğretmen adaylarının soru yazma teknikleri dersinde geliştirdikleri ve ortaokul 8. sınıf öğrencilerine uyguladıkları bir test %70 doğru yanıt yöntemini örneklemek amacıyla kullanılmıştır. Kırk üç kişiden oluşan 4. sınıfta öğrenim gören matematik öğretmen adaylarıyla sekiz çalışma grubu oluşturulmuş, 8. sınıf öğrencileri için testler geliştirilmiş ve uygulanmıştır. Her bir test 100 kişilik farklı 8. sınıf öğrencilerine uygulanmıştır. Bu testler içerisinde madde parametreleri ve öğrenci puan dağılımları iyi olan bir test, puanlara anlam kazandırma yöntemini örneklemek amacıyla seçilmiştir. Bu çalışmada seçilen test 6 öğretmen adayı tarafından geliştirilmiş, uygulanmış, puanlanmış ve puanlara anlam kazandırma çalışması yapılmıştır.

Veri Toplama Aracı

Soru yazma teknikleri dersi kapsamında ilköğretim matematik öğretmeni adayları 20'şer soruluk testler geliştirmişlerdir. Bu ders kapsamında kendilerine test geliştirme basamakları ve soru yazarken dikkat etmeleri gereken hususlar aktarılmıştır. Ayrıca, öğretim programlarına daha paralel olduğu için ağırlıklı olarak TIMSS soruları olmak üzere, PISA soruları gibi geniş ölçekli sınav soru örnekleri öğretmen adayları ile incelenmiştir. Ek olarak, akıl yürütme becerisi ile ilişkili olarak rutin olmayan problemler ile ilgili esaslar paylaşılmış ve ilgili örnekler tartışılmıştır. Öğretmen adaylarından rutin olmayan problemler yazmaları istenmiş, bu örnekler sınıf olarak değerlendirilerek bu konuda ortak bir anlayış geliştirilmiştir. Matematik soruları geliştirilirken üç kategorili TIMSS matematiksel düşünme süreci boyutları temel alınmıştır. İlk bilişsel alan basamağı olan bilme basamağında olgular kavramlar ve süreçler incelenmektedir. İkinci basamak uygulama basamağıdır. Bu basamakta bilgiyi ve kavramsal anlamayı uygulayabilmek için gerekli olan yeteneğe odaklanılır. Üçüncü ve son basamak akıl yürütme basamağıdır. Bu basamakta bilindik olmayan, karmaşık içeriğe sahip ve çok aşamalı problemleri başarmak için rutin problemlerin ötesine geçilmektedir (Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009).

Her çalışma grubu ilk olarak TIMSS matematik düşünme süreçleri olan bilme, uygulama ve akıl yürütme becerilerini ve 8. sınıf birinci üniteye yer alan çarpanlar ve katlar, üslü ifadeler ve kareköklü ifadeler konularını temel alarak belirtke tablolarını hazırlamışlardır. Konu bazında ise kazanım ve ders saati ile orantılı bir dağılım yapılmıştır. Belirtke tabloları incelenip onaylandıktan sonra, soru hazırlama sürecine geçilmiştir. Hazırlanan sorular ilgili kazanımı en iyi örneklendirebilecek soru formatına uygun olarak çoktan seçmeli veya açık uçlu sorulardan oluşmaktadır. Sorularla eş zamanlı olarak, özellikle açık uçlu sorular için, puanlama rehberleri de hazırlanmıştır. Puanlama rehberlerinin sorularla eş zamanlı hazırlanması sayesinde, soru yazarları öğrencilerden gelebilecek tam doğru, kısmi doğru ve yanlış yanıtları önceden tanımlayıp örneklendirdiği için yazdıkları sorulara daha eleştirel gözle bakabilmekte ve hatalardan arınık hale getirebilmektedirler.

Hazırlanan sorular farklı çalışma grubu ve makalenin yazarlarından biri tarafından bilimsel doğruluk ve kazanım düşünme süreci uyumu bakımından incelenmiş ve düzeltmeler verilmiştir. Bu düzeltmelerden sonra, başka bir çalışma grubu ve makalenin yazarlarından biri tarafından yeniden incelenerek sorulara son hali verilmiştir. Ardından sorular öğretmen adaylarının staj yaptıkları ortaokullardaki 100 kişilik 8. sınıf öğrenci grubuna uygulanmıştır. Elde edilen öğrenci yanıtları ITEMAN 4.4 (ASC, 2017) programı kullanılarak analiz edilmiş, eğer varsa, problemleri sorular öğretmen adayları tarafından tespit edilmiştir.

Puanlara anlam kazandırmak amacıyla geliştirilen ve bu çalışma için seçilen testin ITEMAN 4.4 programı kullanılarak elde edilen klasik test kuramı madde parametreleri Tablo 1'de verilmektedir. Geliştirilen test dokuz çoktan seçmeli ve on bir yanıtı sınırlandırılmış açık uçlu sorudan oluşmaktadır. Soruların %35'i bilme, %45'i uygulama ve %20'si akıl yürütme sorusudur. Düşünme süreçlerini örneklendiren birer soru Ek-1'de verilmektedir. Cronbach alfa ile ölçülen güvenilirlik değeri 0.87 olarak tespit edilmiştir. Bu değer George ve Mallery (2003) belirttikleri

kriterlere göre “iyi” derecede bir güvenilirliğe işaret etmektedir. Maddelerin ortalama güçlüğü 0.69, ortalama ayırt ediciliği ise 0.44 olarak hesaplanmıştır. Sadece 1. sorunun ayırt edicilik değeri düşüktür ($r=0.15$). Bu değer sorunun neredeyse herkes tarafından doğru yanıtlandığı için ($p=0.92$) ortaya çıktığı düşünülmektedir. Bu sebeple sorunlu bir soru olarak değerlendirilmemiştir. Ek-2’de verilen toplam puan gruplarına göre çizilmiş doğru yanıtlama yüzdeleri grafiğinde ise öğrencilerin toplam puanı arttıkça bu soruyu doğru yanıtlama yüzdelerinin de arttığı görülmektedir.

Tablo 1. Geliştirilen testin madde parametreleri

Soru	Soru Türü	Düşünme Süreci	Madde Güçlüğü	Ayırt edicilik
1	Çoktan Seçmeli	Bilme	0.92	0.15
2	Açık Uçlu	Uygulama	0.57	0.45
3	Çoktan Seçmeli	Uygulama	0.91	0.32
4	Çoktan Seçmeli	Bilme	0.85	0.35
5	Çoktan Seçmeli	Bilme	0.94	0.56
6	Çoktan Seçmeli	Uygulama	0.64	0.34
7	Açık Uçlu	Akıl Yürütme	0.53	0.48
8	Çoktan Seçmeli	Uygulama	0.65	0.45
9	Açık Uçlu	Bilme	0.91	0.47
10	Çoktan Seçmeli	Uygulama	0.77	0.56
11	Açık Uçlu	Bilme	0.67	0.50
12	Açık Uçlu	Bilme	0.84	0.50
13	Açık Uçlu	Akıl Yürütme	0.43	0.36
14	Çoktan Seçmeli	Uygulama	0.66	0.33
15	Açık Uçlu	Uygulama	0.62	0.52
16	Açık Uçlu	Akıl Yürütme	0.19	0.37
17	Açık Uçlu	Bilme	0.76	0.56
18	Açık Uçlu	Uygulama	0.74	0.59
19	Açık Uçlu	Akıl Yürütme	0.36	0.45
20	Çoktan Seçmeli	Uygulama	0.92	0.44

Veri Analizi

Bu kısımda yeterlik tanımları temel alınarak puanlara anlam kazandırmanın öğretmen adayları tarafından da yapılabilmesine olanak sağlayıp sağlamadığı araştırılan %70 doğru yanıt yöntemi ile ilgili detaylar paylaşılmaktadır. Ardından, madde tepki kuramı kullanılarak madde haritalama yöntemi ile yeterliklerin nasıl belirlendiği aktarılmaktadır.

Yüzde Yetmiş Başarı Oranı Baz Alınarak Yeterlik Tanımlarının Belirlenmesi

Birinci araştırma sorusunu yanıtlamak amacı ile öğretmen adayları geliştirdikleri testleri kullanarak yeterlik tanımları oluşturmuşlardır. Yeterlik tanımlarını oluşturmak için öğretmen adayları %70 doğru yanıt yöntemini kullanmışlardır. Bu yöntemin detayları bu kısımda açıklanmaktadır.

Bu çalışmada incelenen yöntemin başarılı olabilmesi ve puanlara anlam kazandırabilmek için ön koşul, test geliştirilirken test geliştirme basamaklarının (Crocker & Algina, 1986) titizlikle uygulanmış olmasıdır. Öncelikle, konu ve düşünme süreçlerini iyi örneklendiren ve testin amacına uygun bir belirtke tablosu hazırlanmalıdır. Testin iyi bir şekilde planlanması elde edilecek çıktılar açısından da oldukça önemlidir. Sorular hazırlanırken her sorunun ne ölçtüğünün tanımlanması ve belirtke tablosundaki hedef düşünme süreci ile uyumlu olması gerekmektedir. Her bir soru sadece bir düşünme sürecini ölçmelidir. Ayrıca, sorular ölçülmek istenen beceri dışındaki faktörlerden arınık olmalıdır.

Bu ön koşullar sağlandıktan sonra %70 doğru yanıt yöntemi kullanılarak yeterlik tanımlarının belirlenmesi ve puanlara anlam kazandırılması için sırasıyla şu basamaklar uygulanmalıdır. Bu basamaklar Berberoğlu (2006) tarafından teorik yapısı aktarılan yöntem

temel alınarak, çalışmanın yazarlarının uygulama deneyimleri eklenerek genişletilmiş halidir. Her basamakta yapılacak işlemlerin sonunda elde edilmesi beklenen sonuçların bir örneği Tablo 2’de verilmektedir.

- 1) Her öğrencinin sınavdan aldığı toplam puan yüzlük sisteme dönüştürülmelidir.
- 2) Öğrenci puanları sıralanarak, 5 kategori olacak şekilde gruplandırılmalıdır (1-20; 21-40; 41-60; 61-80; 81-100).
- 3) Her puan aralığında kaç öğrenci olduğu belirlenmeli ve üçüncü sütuna yazılmalıdır.
- 4) Her gruptaki öğrenci sayısının %70’i hesaplanıp en yakın tam sayıya yuvarlanarak dördüncü sütuna yazılmalıdır. Bu sayı ilgili puan aralığındaki öğrencilerin yüzde yetmişinin veya daha fazlasının hangi soruları doğru yanıtladığını belirlemede kullanılacaktır.
- 5) Ardından, 1-20 puan kategorisinde yer alan öğrencilerin en az %70’inin hangi soruları doğru yanıtladığı belirlenerek beşinci sütuna yazılmalıdır. Diğer puan kategorileri için de sırasıyla bu işlem yapılmalıdır.
- 6) Genel olarak Tablo 2’de verilen örnekteki gibi genel bir hiyerarşik yapının oluşması beklenmektedir. Örneğin, toplam puanı 1-20 puan arasında olan öğrencilerin (6 öğrenci) en az yüzde yetmişi (4 öğrenci veya daha fazlası) 5 ve 13. soruyu doğru yanıtlamışlardır. 21-40 puan aralığındaki öğrencilerin en az yüzde yetmişi ise 5 ve 13. soru yanı sıra 9 ve 11. soruyu da doğru yanıtlamışlardır. Bu yapıyı bozan, diğer bir deyişle bir alt grupta doğru yanıtlanırken, bir üst grupta doğru yanıtlanamayan sorular olabilir. Ancak, iyi hazırlanmış bir teste bu tip soruların az sayıda olması beklenir.
- 7) Sorular sıralandıktan sonra her kategoride ilk kez yer alan soruların ölçtüğü becerilerin sözel tanımlayıcılarla betimlenmesi sonucunda yeterlik tanımları oluşturulur. Bu aşamada öğretmenlerin grup halinde çalışarak, soruları hazırlarken hedefledikleri tanımlarından yararlanmaları faydalı olacaktır. Örneğin, 5 ve 13. sorunun ölçtüğü beceri “tek aşamalı dört işlemleri yapar” olarak tanımlanırken, 9 ve 11. soruların ölçtüğü beceri “birden fazla basamağı olan dört işlemleri yapar” olarak tanımlanmıştır.
- 8) Hazırlanan yeterlik tanımları sayesinde öğrencilerin becerileri hakkında somut bilgilere sahip olunur. Örneğin, 1-20 puan aralığında yer alan öğrenciler büyük ihtimalle “tek aşamalı dört işlemleri yapma” becerisine sahip iken, 21-40 puan aralığında yer alan öğrenciler ise büyük ihtimalle “tek aşamalı dört işlemleri yapma” becerisine ek olarak “birden fazla basamağı olan dört işlemleri yapma” becerisine de sahiptir.
- 9) Bir öğrencinin yer aldığı puan aralığının üstünde yer alan becerilere ise bu öğrencinin büyük olasılıkla ulaşamadığı düşünülmektedir. Örneğin, 21-40 puan aralığında yer alan bir öğrencinin büyük ihtimalle “temel problemleri çözer” ve daha üst düzeylerde yer alan becerilerde eksikleri vardır.

Tablo 2. %70 Doğru yanıt yöntemi kullanılarak oluşturulmuş yeterlik tanımları örneği

Düzyey	Puan Aralığı	Gruptaki Kişi Sayısı	Grupta yer alanların %70'i	Grubun en az %70'inin yapabildiği sorular	Yeterlik Tanımları
1	1- 20	6	4	s5, s13	Tek aşamalı dört işlemleri yapar.
2	21- 40	12	8	s5, s13 s9, s11	Birden fazla basamağı olan dört işlemleri yapar.
3	41- 60	14	10	s5, s13 s9, s11, s1, s2, s16,	Temel problemleri çözer.
4	61- 80	13	9	s5, s13 s9, s11, s1, s2, s16, s3, s4, s6, s7, s8, s12, s18	Birden fazla işlem gerektiren temel problemleri çözer.
5	81- 100	5	3	s5, s13 s9, s11, s1, s2, s16, s3, s4, s6, s7, s8, s12, s18, s10, s14, s15, s17, s19, s20	Rutin olmayan problemleri çözer.

Madde Tepki Kuramı Kullanılarak Yeterliklerin Belirlenmesi

İkinci araştırma sorusunu kapsamında öğretmen adaylarının puanlara anlam kazandırmak için oluşturdukları yeterlik tanımları tablosundaki (Tablo 3) soru grupları ile madde tepki kuramı temel alınarak elde edilen soru gruplarının benzerliği incelenmiştir. Madde tepki kuramına dayalı olarak madde parametrelerini elde etmek için birçok yazılım mevcuttur. Bu çalışmada, öğrencilerin toplam ham puanını, öğrenci yetenek kestirimini ve madde güçlük değerlerini Wright haritasında vererek madde haritalama yöntemine uygun şekilde gösteren ConstructMap 4.6 (Kennedy, Wilson, Draney, Tutuncuyan, & Vorp, 2010) programı tercih edilmiştir. ConstructMap programı aynı anda hem çoktan seçmeli soruları hem de açık uçlu soruları analiz edebilmektedir. Program madde tepki kuramını temel alarak, 0-1 şeklinde puanlanan soruları Rasch modeli ile, 0-1-2 şeklinde puanlanan soruları da kısmi puan modeli (partial credit model) ile analiz etmektedir. Madde tepki kuramında madde güçlüğü tanımı kapsamında, her bir soru ile aynı düzeyde olan öğrencilerin o soruyu %50 olasılıkla doğru yanıtlanması beklenmektedir.

Çalışmada, ConstructMap programı ile Wright haritası elde edilmiştir. Ardından elde edilen Wright haritasındaki soru grupları incelenmiştir. Öğretmen adaylarının %70 doğru yanıt yöntemi ile elde ettikleri yeterlik tanımlarını belirlemede kullanılan soru grupları ile karşılaştırma yapabilmek için toplam puan 1-20, 21-40, 41-60, 61-80, 81-100 olacak şekilde 5 gruba ayrılmıştır. Her grupta yer alan soru grupları raporlanmıştır.

BULGULAR

%70 doğru yanıt yöntemiyle elde edilen soru sınıflarının bilişsel süreçlere uygunluğuna ilişkin bulgular

Öğretmen adayları geliştirdikleri ve uyguladıkları test sonuçlarını kullanarak %70 doğru yanıt yöntemi ile soru gruplarını belirlemişlerdir. Ardından soru gruplarını inceleyerek yeterli tanımlarını elde etmişlerdir. Bu sürece çalışmanın yazarları da destek vermiştir. Elde edilen sonuçlar Tablo 3’de verilmektedir.

Elde edilen yeterli tanım tablosunda yer alan (Tablo 3) soru grupları incelendiğinde hiyerarşik bir yapı olduğu görülmektedir. Sadece 1 ve 3. soru bu hiyerarşik yapıyı biraz bozmaktadır. Bu sorular birinci düzeydeki öğrencilerin en az %70’i tarafından doğru yanıtlanırken, ikinci düzeydeki öğrencilerin en az %70’i tarafından soru 1, üçüncü düzeydeki öğrencilerin en az %70’i tarafından soru 3 doğru yanıtlanmıştır. Bu iki soru dördüncü ve beşinci düzeyden itibaren yine öğrencilerin en az %70’i tarafından doğru yanıtlanmışlardır. Bunun dışında 16. soru tüm gruba zor gelmiş ($p=0.19$) ve beşinci düzeydeki öğrencilerin %70’i tarafından bile doğru yanıtlanamamıştır. Diğer sorular için hiyerarşik bir yapı ortaya çıkmıştır. İkinci düzeydeki öğrenciler birinci düzeydeki öğrencilerden farklı olarak 4, 5, 9 ve 20. soruları; üçüncü düzeydeki öğrenciler ikinci düzeydeki öğrencilerden farklı olarak 10, 12 ve 17. soruları; dördüncü düzeydeki öğrenciler üçüncü düzeydeki öğrencilerden farklı olarak 8, 11, 15 ve 18. soruları; beşinci düzeydeki öğrenciler dördüncü düzeydeki öğrencilerden farklı olarak 2, 6, 7, 13, 14 ve 19. soruları doğru yanıtlanmışlardır. Bu bakımdan %70 doğru yanıtlanma yöntemine göre elde edilen sonuçlar teorik olarak tanımlandığı gibi hiyerarşik yapıda birer soru grubu oluşturabilmektedir.

Soru grupları kullanılarak elde edilen yeterli tanımları incelendiğinde de testin geliştirilmesinde temel alınan ve TIMSS matematiksel düşünme becerileri olan bilme, uygulama ve akıl yürütme becerilerinden oluşan hiyerarşik yapı gözlemlenebilmektedir. Birinci ve ikinci düzeydeki yeterli tanımlarının genel olarak bilme becerisi ile ilgili olduğu söylenebilir. Bu düzeylerde öğrencilerin ilgili matematiksel konuları bilme ve tanımasına yönelik sorular bir araya gelmiştir. Üçüncü ve dördüncü düzeydeki yeterli tanımlarının ise ağırlıklı olarak uygulama becerisi ile ilgili olduğu görülmektedir. Bu düzeylerde öğrencilerin işlem yapmasını ve rutin problem çözmesini gerektiren sorular bir grup oluşturmuştur. Beşinci düzeyde ise akıl yürütme becerisi ile ilgili olarak rutin ve rutin olmayan problemlerin bir araya geldiği görülmektedir. Özetle, %70 doğru yanıt yöntemine göre elde edilen sonuçlar bilme, uygulama ve akıl yürütme becerilerinden oluşan hiyerarşik yapı ile uyumludur.

Burada elde edilen yeterli tanımları öğrencilerin sahip oldukları sınav puanlarına anlam kazandırmaları bakımından oldukça önemlidir. Geliştirilen bu test özelinde, örneğin 50 puan alan ve bu sebeple üçüncü düzeyde yer alan bir öğrenci, çok büyük olasılıkla birinci ve ikinci düzeyde yer alan beceriler olan tam sayıların tam sayı kuvvetlerini bilmekte, çok büyük ve çok küçük sayıların bilimsel gösterimini bilmekte, tam kare doğal sayıları tanımakta, bir pozitif tam sayının asal çarpanlarını bulabilmekte ve tamkare olmayan kareköklü bir sayının hangi iki doğal sayı arasında olduğunu belirleyebilmektedir. Bu öğrenci yer aldığı üçüncü düzeydeki beceriler olan üslü ifadelerle ilgili işlemleri yapabilmek, belli iki sayı arasındaki tam kare doğal sayıları listeleyebilmek, kareköklü bir ifadeyi $a\sqrt{b}$ şeklinde yazabilmek ve $a\sqrt{b}$ şeklindeki ifadede katsayıyı kök içine alabilmek becerilerine de büyük olasılıkla sahiptir. Bir öğrencinin sahip olduğu becerileri bilebilmek oldukça önemlidir. Bu becerileri temel alarak eksik olan becerilerin giderilmesine odaklanılmalıdır. 50 puan alan öğrencinin, dördüncü ve beşinci düzeyde tanımlanan becerilerde eksikleri vardır. Bu öğrenci rutin ve rutin olmayan problem çözme becerilerine çok büyük olasılıkla sahip değildir.

Tablo 3. Öğretmen adayları tarafından oluşturulmuş yeterlik tanımları

Düzye	Puan Aralığı	Gruptaki Kişi Sayısı	Grupta yer alanların %70'i	Grubun %70'inin yapabildiği sorular	Bilişsel süreçler	Yeterlik Tanımları
1	1-20	5	4	S1, S3	Bilme: S1 Uygulama:S3	<ul style="list-style-type: none">Tam sayıların tam sayı kuvvetlerini bilir.
2	21-40	10	7	S1 S4, S5, S9, S20	Bilme: S1, S4, S5, S9 Uygulama:S20	<ul style="list-style-type: none">Çok büyük ve çok küçük sayıların bilimsel gösterimini bilir.Tam kare doğal sayıları tanıır.Bir pozitif tam sayının asal çarpanlarını bulur.Tamkare olmayan kareköklü bir sayının hangi iki doğal sayı arasında olduğunu belirler.
3	41-60	15	11	S3 S4, S5, S9, S20 S10, S12, S17	Bilme: S1, S4, S5, S9, S12, S17 Uygulama:S10, S20	<ul style="list-style-type: none">Üslü ifadelerle ilgili işlemleri yapar.Belli iki sayı arasındaki tam kare doğal sayıları listeler.Kareköklü bir ifadeyi $a\sqrt{b}$ şeklinde yazar ve $a\sqrt{b}$ şeklindeki ifadede katsayıyı kök içine alır.
4	61-80	47	33	S1, S3 S4, S5, S9, S20 S10, S12, S17 S8, S11, S15, S18	Bilme: S1, S4, S5, S9, S11, S12, S17 Uygulama:S3, S8, S10, S15, S18, S20	<ul style="list-style-type: none">Sayıların ondalık gösterimlerini 10'un tam sayı kuvvetlerini kullanarak çözümler.Kareköklü ifadeleri sıralar.Ondalık ifadelerin karekökleri ile ilgili rutin problemleri çözer.Tamkare olmayan kareköklü bir sayının hangi iki doğal sayı arasında olduğu ile ilgili rutin problemleri çözer.
5	81-100	23	16	S1, S3 S4, S5, S9, S20 S10, S12, S17 S8, S11, S15, S18 S2, S6, S7, S13, S14, S19	Bilme: S1, S4, S5, S9, S11, S12, S17 Uygulama:S2, S3, S6, S8, S10, S14, S15, S18, S20 Akıl yürütme: S7, S13, S19	<ul style="list-style-type: none">Çok büyük ve çok küçük sayıların bilimsel gösterimi ile ilgili rutin problemleri çözer.Tam kare doğal sayılar ile ilgili rutin problemleri çözer.Verilen iki doğal sayının aralarında asal olup olmadığı ile ilgili rutin olmayan problemleri çözer.İki doğal sayının en büyük ortak bölenini (EBOB) ve en küçük ortak katını (EKOK) hesaplama ile ilgili rutin olmayan problemleri çözer.

* Soru 16 hiçbir grup tarafından %70 oranında doğru yanıtlanmamıştır.

%70 Doğru Yanıt Yöntemiyle Elde Edilen Soru Gruplarının Madde Haritalama Yöntemi ile Karşılaştırılması

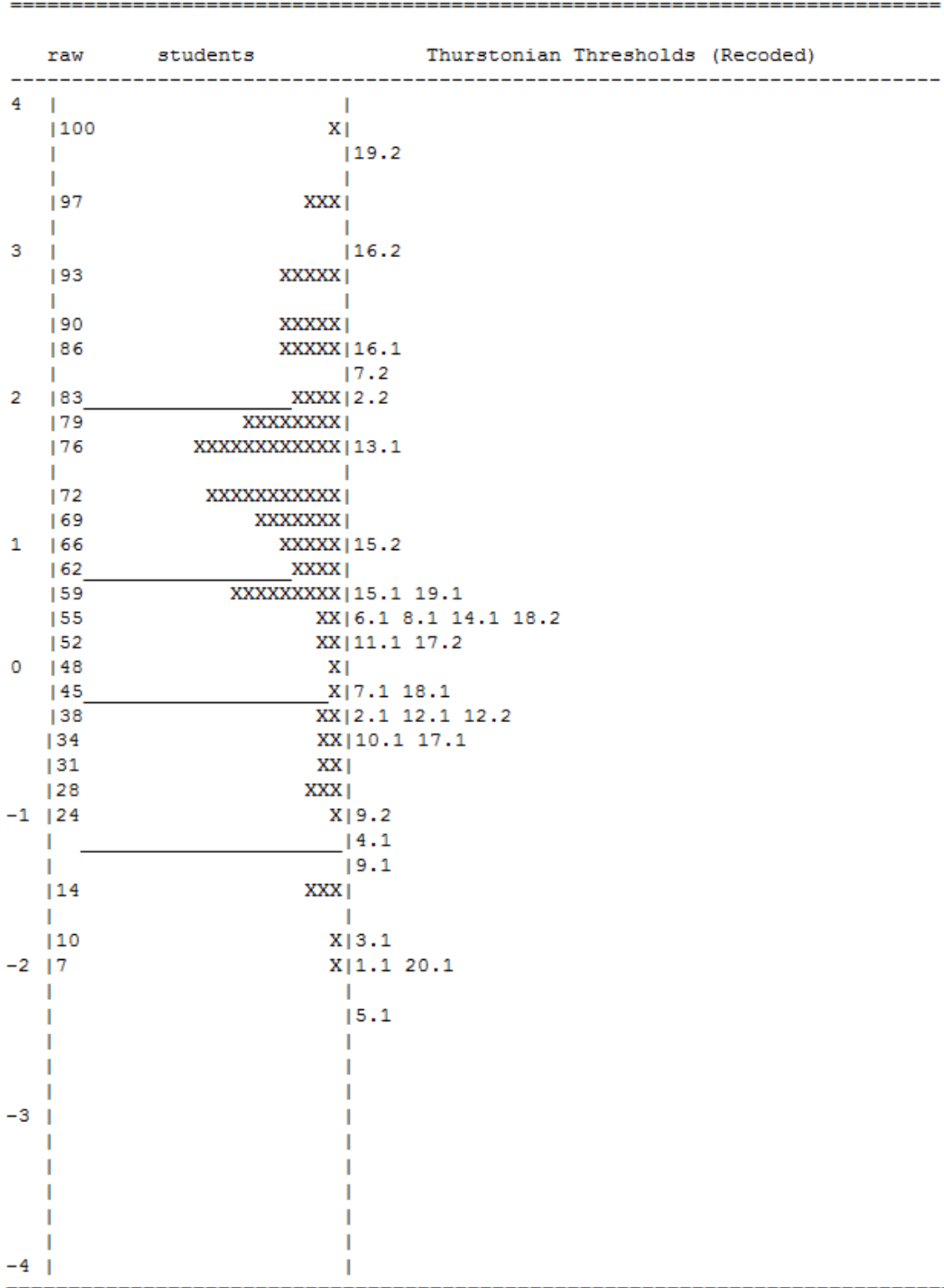
Bu çalışmanın ikinci araştırma sorusu öğretmen adayları tarafından %70 doğru yanıt yöntemi temel alınarak oluşturulan soru grupları ile madde tepki kuramı Wright haritası ile elde edilen soru gruplarının uyumunu karşılaştırmaktır. ConstructMap programı kullanılarak elde edilen Wright haritası Şekil 1’de verilmektedir. Wright haritasının en solunda -4 ile +4 arasında değişen öğrenci yetenek kestirim aralıkları verilmektedir. Bu sütunun sağında (raw sütunu) öğrencilerin aldıkları ham puanlar görülmektedir. Bu test için öğrenciler 7 ile 100 puan arasında puanlar elde etmişlerdir. Bir sağ sütunda ise (students sütunu) her puan düzeyindeki öğrenci sayısını temsil eden X işaretleri yer almaktadır. Bu testten alınan puanların dağılımı ve çarpıklık (-1.025) ve basıklık (0.608) katsayıları incelendiğinde sonuçlarının genel olarak sola çarpık bir dağılım gösterdiği görülmektedir. En sağ sütunda ise (Thurstonian Thresholds sütunu) testte yer alan sorular zorluk sıralarına göre yer almaktadır. Bir öğrencinin aynı hizada bulunduğu bir soruyu %50 olasılıkla doğru yanıtlaması beklenirken, kendi yetenek kestiriminin üstünde kalan soruları %50’den daha düşük olasılıkla, kendi yetenek kestiriminin altında kalan soruları ise %50’den daha yüksek olasılıkla doğru yanıtlaması beklenmektedir. Örneğin, 0-1-2 şeklinde puanlanan 15. sorudan %50 olasılıkla iki puan alabilmek için (15.2) bir öğrencinin yetenek kestirim değerinin 1 (ham puan 66), 0-1-2 şeklinde puanlanan 18. sorudan %50 olasılıkla bir puan alabilmek için (18.1) bir öğrencinin yetenek kestirim değerinin 0 civarı (ham puan 45) olması beklenmektedir. Ayrıca, ham puanları 1–20, 21–40 v.b. şekilde görebilmek adına Şekil 1’e dört adet yatay çizgi eklenmiştir.

Tablo 4’te %70 doğru yanıt yöntemi ile elde edilen soru grupları ile Wright haritasından elde edilen soru grupları karşılaştırılmıştır. Wright haritası daha önce de belirtildiği üzere soruların %50 doğru yanıtlanma olasılığını temel almaktadır. Bu sebeple, Wright haritası yöntemindeki belli bir düzeyde yer alan bir sorunun %70 doğru yüzdesi yönteminde aynı düzeyde veya bir üst düzeyde yer alması beklenmektedir. Soru 6 ve 14 dışında tüm sorular beklendiği gibi birbirine yakın düzeylerde yer almışlardır. Dokuz soru tamamen aynı düzeyde yer alırken, dokuz soru da bir üst düzeyde yer almaktadır. Hesaplama yöntemlerinin birinde %50 oranı, diğerinde %70 oranı baz alınmasına rağmen soruların düzeylere dağılımları arasında benzerlik olduğu görülmektedir.

Tablo 4. Soru Gruplarının Karşılaştırılması

Düzyen	%70 Doğru Yanıt Yöntemi	Wright Haritası Yöntemi	Aynı düzeyde bulunan sorular	Bir üst düzeyde yer alan sorular
1	S1, S3	S1, S3, S5, S20	S1, S3	S5, S20
2	S4, S5, S9, S20	S4, S9, S10, S12	S4, S9	S10, S12
3	S10, S12, S17	S6 , S8, S11, S14 , S17, S18	S17	S8, S11, S18
4	S8, S11, S15, S18	S13, S15	S15	S13
5	S2, S6 , S7, S13, S14 , S19	S2, S7, S16, S19	S2, S7, S19	S16

Wright Map (EAP) Variable: Construct 1
 IRT Categories
 Map of person estimates and response model parameter estimates



Each X represents 1 student, each row is 0,170 logits

Şekil 1. Geliştirilen teste ait Wright haritası

TARTIŞMA ve SONUÇ

Bu çalışmanın genel amacı, öğretmen adaylarının geliştirdikleri bir matematik testinin uygulama sonuçlarını kullanarak, test sonuçlarından elde edilen puanlara yeterli tanımları bazında anlam kazandırılıp kazandırılmayacağını incelemek ve %70 doğru yanıt yönteminin kullanılabilirliğini değerlendirmektir. Bu genel değerlendirmeyi yapabilmek için iki ölçüt kullanılmıştır. Birinci ölçüt, %70 doğru yanıt yöntemi ile oluşan soru grupları arasında gözlenen hiyerarşik yapının soruların hazırlanmasında temel alınan hiyerarşik yapıdaki matematiksel düşünme süreçleri ile uyumudur. İkinci ölçüt ise %70 doğru yanıt yöntemi ile oluşan soru gruplarının test geliştirme uzmanları tarafından da yaygın olarak kullanılan madde haritalama yönteminden elde edilen soru grupları ile tutarlılık göstermesidir. Çalışmada elde edilen bulgular %70 doğru yanıt yöntemi kullanılarak hiyerarşik bir yapıda soruların gruplandığını ve bu gruplamanın Wright haritası ile belli ölçüde benzerlik gösterdiğine işaret etmektedir. Bu çalışma kapsamında elde edilen bulgular %70 doğru yanıt yöntemi kullanılarak öğrencilere yeterli tanımları bazında anlamlı geri bildirim verilebileceğini ve puanlara anlam kazandırılabilceğini göstermektedir.

Türkiye'deki öğretmenler sınıflarında elde ettikleri ölçme ve değerlendirme sonuçlarını öğrencilere dönüt vermek amacı ile nadiren kullanmaktadırlar (Demirtaşlı, 2009). Bunun pek çok sebebi olabilir. Öğretmenler, sınav sonuçları ile anlamlı geri bildirim vermeleri gerektiğini bilmiyor olabilirler. Üniversite yaşantıları da dahil olmak üzere, bunun örneklerini görmemiş ve bu yönde eğitim almamış olabilirler. Gelbal ve Kelecioğlu (2007) üniversitelerde yer alan öğretmen yetiştirme programlarında ölçme ve değerlendirmeye yeteri kadar zaman ayrılmadığı için öğretmen adaylarının ölçme ve değerlendirmenin amaçları ile ilgili tekniklerde önemli eksikleri olduğunu belirtmişlerdir. Bu sebeple yaptıkları sınavları sadece birer not ve sıralama ile raporluyor olabilirler. Anlamlı geri bildirim vermek gerektiğini düşünen öğretmenler de bunu nasıl yapacaklarını bilmiyor veya bu konuda uzman yardımı alamıyor olabilirler. Bazı özel okullar dışında kalan okullarda bu hizmeti sağlayacak ölçme ve değerlendirme birimleri bulunmayabilir. Öğretmenlerin de kendi iş yükleri arasında ayrıca sınav sonuçlarını analiz etmeyi öğrenip, madde tepki kuramı gibi teknikleri öğrendikleri sıklıkla görülen bir durum değildir. Özellikle, ileri seviyede madde analizi yapmayı bilmeyen veya böyle bir analizi yapmaya imkân olmayan öğretmenler, sahip oldukları temel bilgi ve basit hesaplamalar ile yeterlik tanımlarını oluşturarak puanlara anlam kazandırmayı başarabilirlerse yaptıkları ölçme ve değerlendirmenin esas amacına daha uygun olacağı düşünülmektedir. Yapılan ölçme ve değerlendirme sonuçlarını kullanarak öğrencilere anlamlı geri bildirim sağlanması oldukça önemlidir (Berberoğlu, 2006; Berberoğlu, 2009; Demirtaşlı, 2017). İdeal olan, öğretmenlerin zümreler halinde bir ölçme ve değerlendirme uzmanının desteğini de alarak sınavlarını hazırlamaları ve sınavlardan sonra öğrencilerine yeterlik tanımlarını da içeren sonuçları iletebilmeleridir. Bu sayede öğretmenler de sınıflarını daha yakından tanıyabilecek ve öğrencileri için daha etkili planlamalar yapabilecektir. Okullarda zümrelere destek olacak, ileri seviye madde analizi tekniklerine hakim ölçme ve değerlendirme uzmanlarının da istisnai olarak var olduğu düşünülürse, öğretmenlerin sınav sonuçlarına anlam kazandırabilmek için kolay uygulanabilir ve ana hatları ile de olsa yeterlik tanımlarını elde etmelerini sağlayacak bir yöntem ihtiyaçları vardır. Bu çalışmada detayları paylaşılan, gerçek öğrenci verisi ile uygulanabilirliği değerlendirilen ve madde tepki kuramı kullanılarak sonuçların benzerliği incelenen %70 doğru yanıt yöntemi öğretmenler için puanlara anlam kazandırma amacı ile başlangıç seviyesinde kullanılabilir bir yöntemdir. Bu yöntemi kullanan öğretmenler elde ettikleri ölçme ve değerlendirme sonuçlarını öğrencilere dönüt vermek amacı ile daha sık kullanabileceklerdir.

%70 doğru yanıt yöntemi ile elde edilen yeterlik tanımları sayesinde öğrencilerin sahip oldukları ve geliştirmeleri gereken bilgi ve beceriler ortaya çıkabilmektedir. Öğretmenler öğrencilerine “üslü ifadelerle ilgili işlemleri yapabilmek, belli iki sayı arasındaki tam kare doğal sayıları listeleyebilmek, kareköklü bir ifadeyi $a\sqrt{b}$ şeklinde yazabilmek ve $a\sqrt{b}$ şeklindeki ifadeye katsayıyı kök içine alabilmek becerilerine sahipsin ama rutin ve rutin olmayan problem çözme becerilerine çok büyük olasılıkla eksiklerin var” gibi somut dönütler verebilmelidir. Bu sayede bir

öğretmen genel hatları ile bir öğrencisine ve velisine öğrencinin sahip oldukları becerileri somut olarak tanımlayabilecektir. Elinde öğrencileri hakkında bu tip somut bilgileri olmayan öğretmenler öğrencilerine veya velilerine sadece öğrencinin sınavdan aldığı notlardan oluşan geri bildirimlerde bulunabilmektedir.

Somut yeterlik tanımlarına sahip olunmasının diğer bir faydası da öğrencileri sıralamaya dayanan bağıl (norm dayanaklı) değerlendirmenin kullanımının da giderek azalabilecek olmasıdır. Sıralamaya dayalı sonuçlarda sınıflarda hep belirli öğrenciler en önlere yer almakta ve sıralamada gerilerde yer alan öğrencilerde giderek yaygınlaşan bir “başaramıyorum” algısı oluşmaktadır. Bir öğrencinin başarabildiklerini bilmesi özgüveni bakımından oldukça önemlidir. Özgüven ve başarı arasında bir ilişki olduğu da bilinmektedir (Abu-Hilal, 2000; Marsh, 1986; Shen & Tam, 2008). Öğrenciye sahip olduğu beceriler iletildikten sonra, geliştirmesi gereken yönler de yeterlik tanımları sayesinde bilgi ve beceri bazında iletilebilecektir. Temel matematiksel bilgilere sahip, belli işlemleri yapabilen ama problem çözmede zorlanan bir öğrencinin artık saatlerce sadece işlem gerektiren etkinlik veya sorular ile zaman kaybetmemesi gerektiği açıktır. Bunun yanında, temel bilgilere bile ulaşamamış bir öğrencinin rutin olmayan problemleri çözebilmesi için biraz daha zamana ihtiyacı vardır. Bu bilgiler, öğretmenin doğru planlama yapabilmesine olanak sağlayacaktır. Benzer şekilde yeterlik tanımları sayesinde öğretmen sınıfındaki genel düzey hakkında da bilgi sahibi olabilecektir. Derslerde öğrenciler arasında başarı farkları ortaya çıkmaktadır. Örneğin 21-40 puan arasında başarı gösteren öğrenciler ile 61-80 puan arasında başarı gösteren öğrenciler arasındaki beceri farklılıklarının daha somut bilgiler ışığında ortaya konulmasına ihtiyaç vardır. Bu sayede öğretmenler tüm sınıfa aynı ödevleri vermek yerine bireyselleşmiş, kapsamı benzer ama farklı ağırlıklarda ödevler hazırlayabilirler. Kişileştirilmiş ödevlerin başarı ile ilişkili olduğuna dair bulgular mevcuttur (Liberatore, 2011).

Elde edilen bu sonuçlar, öğretmen adaylarının veya öğretmenlerin doğru bir yönlendirme ile geliştirdikleri testlerdeki puanlara anlam kazandırabileceklerini göstermektedir. Bu yöntem kesinlikle madde tepki kuramı kullanılarak tecrübeli uzmanlar tarafından ortaya konan yeterlik tanımlarına bir alternatif değildir. Ancak, uzmanlar tarafından oluşturulmuş yeterli tanımları içeren ölçme ve değerlendirme sonuçlarına ulaşamayan öğretmenler için sadece puan raporlamaya göre oldukça değerli bilgiler elde edilmesini sağlayan bir yöntemdir. Bu çalışma Berberoğlu (2006) tarafından teorik altyapısı tanıtılan %70 doğru yanıt yöntemi ile elde edilen bulguların savunulabilir ve kullanılabilir olduğunu göstermiştir.

Sınırlılıklar

Bu çalışma ile ilgili sınırlılıkları da listelemek gerekmektedir. Öncelikli olarak bu çalışmada elde edilen soru gruplama sonuçları öğrenci grubunun başarısına ve kullanılan soruların zorluk derecelerine bağlıdır. Klasik test kuramına benzer şekilde, elde edilen sonuçlar farklı bir grupta veya daha kolay veya daha zor sorulardan oluşan bir test ile farklılık gösterebilecektir. Elde edilen bulgular, ilgili öğrenci grubu için geçerlidir. Çok daha başarılı bir grupta soruların yanıtlanma yüzdeleri değişecektir. Buna bağlı olarak 1-20 puan aralığı gibi düzeylerde hiçbir öğrenci olmayabilir. Bu çalışmada örnek olarak gösterilen test kapsamında sorular orta seviye denilebilecek bir güçlük ortalamasına sahiptirler. Gruba çok zor gelen veya çok basit gelen sınavlarda burada elde edilen hiyerarşik yapılar gözlemlenmeyebilir.

Araştırmanın bir diğer sınırlılığı ise çalışmada kullanılan kişi sayısıdır. Az sayıda öğrenci ile (20-30 civarı) bu yöntem denendiğinde her grupta 4-5 kişi olacağı düşünülürse, soruların yanıtlanma yüzdeleri gruptaki bir iki kişinin doğru yanıtlayıp yanıtlayamamasından kolaylıkla etkilenecektir. Tutarlı sonuçlar için kişi sayısının az olmaması gerekmektedir. Bu çalışmada 100 öğrenci ile iyi sonuçlar elde edilebildiği görülmektedir. Öğrenci sayısının fazla olması tutarlı sonuçlar için gereklidir. Öğrenci sayısının az olduğu durumlarda 5 puan grubu yerine 3 puan grubu kullanılması faydalı olabilir.

Hazırlanan testin temel özelliklere sahip olmaması da bu yöntemin tutarlı sonuçlar vermesinde önemli bir engel olacağı düşünülmektedir. Güvenilirlik katsayısı düşük, bilimsel hatalar bulunan ve ölçülen becerilerin net olmadığı sorulardan oluşan bir test ile puanlara anlam

kazandırabilmek oldukça sıkıntılı olacaktır. Testi hazırlayan öğretmenlerin temel ölçme ve değerlendirme becerilerine sahip olmaları bir önkoşuldur.

Öneriler

Burada belirtilen sınırlılıklar dikkate alındığında %70 doğru yanıt yöntemi ile elde edilen sonuçları etkileyebilecek soru sayısı, puan dağılımları ve soru güclüğü öğrenci becerisi uyumu gibi faktörlerin sonuçlar üzerindeki etkisi incelenmelidir. Ayrıca, bu çalışmada %70 doğru yanıt oranı genel olarak öğrencilerin kazanması beklenen bir değer olarak alınmıştır. Bu konuda yapılacak araştırmalarda, %60 veya %80 gibi yüzdeler baz alındığında sonuçların nasıl değişeceği incelenebilir. Ayrıca bu araştırmadan elde edilen sonuçlar Wright haritasına dayalı sonuçlarla karşılaştırılarak incelenmiştir. İleride yapılacak araştırmalarda %70 doğru yanıt yöntemi, başka yeterlik belirleme yöntemleri ile karşılaştırılabilir.

KAYNAKÇA

- Abu-Hilal, M. M. (2000). A structural model for predicting mathematics achievement: Its relation with anxiety and self-concept in mathematics. *Psychological Reports, 86*(3), 835-847.
- Arıkan, S. (2017). Sınıf içi ölçme sonuçlarına dayalı olarak not verme. N. Çıkrıkçı-Demirtaşlı, (Ed.), *Eğitimde Ölçme ve Değerlendirme İçinde* (4.baskı) (349-377). Ankara: Anı Yayıncılık
- ASC (2017). *ITEMAN 4.4 [computer software]*. Minneapolis, MN.
- Berberoğlu, G. (2006). *Sınıf içi ölçme ve değerlendirme teknikleri*. İstanbul: Morpa Kültür Yayınları
- Berberoğlu, G. (2009). Madde haritalama yöntemi ve CİTO Türkiye öğrenci izleme sistemi (öis) uygulamalarında yeterlik düzeylerinin belirlenmesi. *Cito Eğitim: Kuram ve Uygulama, 3*, 13-24.
- Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston
- Demirtaşlı, N. (2009). Eğitimde niteliği sağlamak: ölçme ve değerlendirme sistemi örneği olarak CİTO Türkiye öğrenci izleme sistemi (ÖİS). *Cito Eğitim: Kuram ve Uygulama, 3*, 25-38.
- Demirtaşlı, N. (2017). Öğrenme, Öğretim ve Değerlendirme Arasındaki İlişkiler. N. Demirtaşlı, (Ed.), *Eğitimde Ölçme ve Değerlendirme İçinde* (4.baskı) (1-27). Ankara: Anı yayıncılık
- Gelbal, S., & Kelecioğlu, H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 33*, 135-145.
- George, D., & Mallery, P. (2003). *SPSS for windows step by step: a simple guide and reference. 11.0 update (4th ed.)*. Boston: Allyn & Bacon.
- Kennedy, C. A., Wilson, M. R., Draney, K., Tutuncuyan, S., & Vorp, R. (2010). *ConstructMap 4.6. [computer software]*. Berkeley, California.
- Liberatore, M. W. (2011). Improved student achievement using personalized online homework. *Chemical Engineering Education, 45*(3), 184-190.
- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23*(1), 129-149.
- Martin, M. O., Mullis, I. V. S., & Hooper, M. (2016). "Methods and procedures in TIMSS 2015." Retrieved on 15 March 2017, at URL: <https://timssandpirls.bc.edu/publications/timss/2015-methods.html>
- MEB (2017). *Akademik becerilerin izlenmesi ve değerlendirilmesi 8. sınıf raporu*. Erişim Adresi http://odsgm.meb.gov.tr/meb_iys_dosyalar/2017_11/30114819_iY-web-v6.pdf
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). "TIMSS 2015 International results in mathematics." Retrieved on 15 March 2017, at URL: <http://timssandpirls.bc.edu/timss2015/international-results>
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

- OECD (2016). *PISA 2015 Results (Volume I): Excellence and equity in education*. Paris: OECD Publishing. doi:10.1787/9789264266490-en
- OECD (2017). *PISA 2015 Technical Report*. Paris: OECD Publishing.
- Özgen Tuncer, Ç. (2009). Cito Türkiye Öğrenci İzleme Sistemi. *Cito Eğitim: Kuram ve Uygulama*, 2, 25-52.
- Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data issues in secondary analysis and reporting. *Educational Researcher*, 39(2), 142-151. doi:10.3102/0013189x10363170
- Shen, C., & Tam, H. P. (2008). The paradoxical relationship between student achievement and self-perception: a cross-national analysis based on three waves of TIMSS data. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 14(1), 87-100.

Ek 1. Bilme, Uygulama ve Akıl Yürütme Düşünme Süreçleri Soru Örnekleri

Bilme Düşünme Süreci

5. Aşağıdaki sayılardan hangisi tam kare bir sayıdır?

- A) 63
- B) 121
- C) 1000
- D) 145

Uygulama Düşünme Süreci

14. Bir fabrikada günde 200 milyar pet şişe üretilmektedir. 60 yılda üretilen pet şişe sayısının bilimsel gösterimi aşağıdakilerden hangisidir?

- A) $43,8.10^{15}$
- B) $4,38.10^{13}$
- C) $43,8.10^{13}$
- D) $4,38.10^{15}$

Akıl Yürütme Düşünme Süreci

19. Bir okuldaki öğrenciler 2 öğrenci daha olsaydı 4'erli ve 5'erli gruplara ayrılabilirdi. Okul mevcudu iki basamaklı bir doğal sayıdır.

Buna göre okul mevcudu kaç olabilir? Listeleyiniz.

Ek 2. Soruların farklı başarı seviyelerinde yanıtlanma yüzdeleri (ITEMAN)

