

**TIMSS SORU KİTAPÇIKLARININ KLASİK TEST TEORİSİ VE MADDE TEPKİ
KURAMINA GÖRE ANALİZLERİ: ÖRNEK BİR UYGULAMA****ANALYSIS OF TIMSS QUESTION BOOKLETS ACCORDING TO CLASSICAL TEST
THEORY AND ITEM RESPONSE THEORY: A SAMPLE APPLICATION**

Bedirhan TEKE¹

¹*Arş. Gör., Kilis 7 Aralık Üniversitesi, Kilisli Muallim Rifat Eğitim Fakültesi, Matematik ve Fen Bilimleri Eğitimi Bölümü, Kilis, TÜRKİYE*

¹*ORCID ID: <https://orcid.org/0000-0002-8565-215X>*

Aybige ARABACI²

²*Bolu Abant İzzet Baysal Üniversitesi, Eğitim Fakültesi, Matematik ve Fen Bilimleri Eğitimi Bölümü, Bolu, TÜRKİYE*

²*ORCID ID: <https://orcid.org/0000-0002-4912-5782>*

ÖZET

Bu çalışma, TIMSS Türkiye sekizinci sınıf öğrencilerine yönelik uygulanan kitapçıkların Klasik Test Teorisine (KTT) ve Madde Tepki Kuramına (MTK) göre analizlerinin yapılması ve varsa uyumsuz soru maddelerinin tespitinin nasıl olması gerektiği hakkında bilgi vermek amacıyla hazırlanmıştır. Bu amaç doğrultusunda, TIMSS 2015¹ Türkiye sekizinci sınıf öğrencilerine yönelik uygulanan kitapçıklardan biri olan Kitapçık 3 örnek olarak seçilmiştir. MTK analizlerinde Rasch Model'in kullanılacak olması, bir soru hariç bu kitapçıkta yer alan soru maddelerinin kodlanmasının 0-1 şeklinde olması ve soru maddelerine ulaşılabilir olması bu kitapçığın seçilme gerekçesidir. Kitapçık 3'ün yapısı incelendiğinde, TIMSS 2011'de uygulanan M04 matematik blok soruları kaynak gösterilerek oluşturulmuş M03 matematik bloğu ile TIMSS 2015 kapsamında yeni soru maddeleriyle hazırlanan M04 matematik bloğunun bir araya getirilmesiyle oluşturulmuştur. Her iki blok soruları Sayı (S), Cebir (C), Geometri (G) ve Veri ve Olasılık (VG) içerik alanına sahiptir. Ancak M03 matematik bloğunda yoğunluk bakımından içerik alanlarının sıralamasının $S > C > G = VG$; M04 matematik bloğunda ise bu sıralamanın $S = C > G > VG$ şeklinde olduğu görülmektedir. Bu çalışmada doküman incelemesi yöntemi kullanılmış ve Kitapçık 3'e ait SPSS veri dosyasının uygunluğu ilk olarak incelenmiştir. Bu inceleme sürecinde veri kaybının olup olmadığı göz önünde bulundurulmuştur. Sonuç olarak, 440 öğrencinin cevap verdiği ve 28 soru maddesinden oluşan Kitapçık 3 veri dosyası elde edilmiştir. Elde edilen veri dosyanın MTK analizlerine uyumlu olup olmadığı birtakım varsayımların test edilmesiyle (model uyumu, tek boyutluluk, madde karakteristik eğrisinin (MKE) monotonik olarak artması ve yerel bağımsızlık) sağlanmıştır. KTT analizlerinde ilk olarak, veri setine ait yüzdellik sıra ve z-puan değişkenleri ayrı ayrı hesaplanmıştır. İkinci olarak, veri setine ait güvenilirlik kat sayıları hem TAP programı hem de test yarılama yöntemiyle ayrı ayrı hesaplanmış ve karşılaştırılmıştır. Üçüncü olarak, veri setinde yer alan maddelerin madde güçlük ve ayırt edicilik indeksleri (Adj Pt Bis²) hesaplanmıştır. Son aşamada ise, veri setinde yer alan maddelerin içerik alanları bazında çeldirici analizleri yapılmıştır. MTK analizlerinde ise R studio programının kullanılmasıyla veri setine ait; eta (madde gücüğü), thresholds eşik değeri (bir sorunun yapılma olasılığının %50 olduğu değer), p-value (anlamlılık değeri), outfit MSQ (ağırlaştırılmamış) - infit MSQ (ağırlaştırılmış) değerleri ile outfit t (ağırlaştırılmamış) - infit t (ağırlaştırılmış) değerleri sırasıyla hesaplanmıştır. Bu hesaplamalarda, geçerliliğinin test edilme

¹ Trends in Mathematics and Science Study, (2015). <https://timssandpirls.bc.edu/timss2015/international-database/>

² Adj Pt Bis değeri, ilgili soru maddesinin dahil edilmeden hesaplanan toplam test puanı üzerinden yapılan işlemdir (Adj Pt Bis value is the operation performed on the total test score calculated without including the relevant question item).

sürecinde Wu & Adams (2007)³'ün çalışmaları temel alınmıştır. Bu kapsamda soru maddesinde yer alan çeldiricilerin ters yönlü çalışması, p değerinin 0,05'ten küçük olması, MSQ değerinin 1'in üstünde olması, t değerinin yüksek çıkması ve toplam test puanı ile madde arasında düşük bir ilişkinin olması maddelerin uyumsuz olarak nitelendirilmesine yol açmaktadır. Analizler sonucunda ayırt edicilik ve güçlük indeksinin düşük olduğu, Wu & Adams (2007)'ün ifade ettiği değerleri sağlamayan yedi soru maddesinin model ile uyumsuz olduğu ve bu sebeple bu soru maddelerinin gözden geçirilmesi gerektiği düşünülmektedir.

Anahtar Kelimeler: TIMSS, Klasik Test Teorisi, Madde Tepki Kuramı, Rasch Model, Madde Analizi

ABSTRACT

This study has been prepared in order to provide information on how to analyze the booklets applied to TIMSS Turkey eighth grade students according to Classical Test Theory (CTT) and Item Response Theory (MTK) and how to identify incompatible question items, if any. For this purpose, Booklet 3, which is one of the booklets applied to the eighth grade students of TIMSS 2015 Turkey, was chosen as an example. The reason for choosing this booklet is that the Rasch Model will be used in IRT analyzes, the coding of the question items in this booklet is 0-1 except for one question, and the question items are accessible. When the structure of the booklet 3 is examined, it was created by combining the M03 math block, which was created with reference to the M04 math block questions applied in TIMSS 2011, and the M04 math block, which was prepared with new question items within the scope of TIMSS 2015. Both block questions have Number (S), Algebra (C), Geometry (G) and Data and Probability (VG) content area. However, in the M03 mathematical block, it is observed that the ranking of content areas in terms of density is $S > C > G = VG$; in the M04 mathematical block, this ranking is $S = C > G > VG$. In this study, the document analysis method was used and the suitability of the SPSS data file belonging to Booklet 3 was examined first. In this review process, it was taken into account whether there was any data loss. As a result, Booklet 3 data file consisting of 28 questions and answered by 440 students was obtained. The compatibility of the obtained data file with the IRT analyzes was ensured by testing some assumptions (model fit, unidimensionality, monotonically increasing item characteristic curve (ICC) and local independence). In the CTT analyzes, firstly, the percentile rank and z-score variables of the data set were calculated separately. Secondly, the reliability coefficients of the data set were calculated with both the TAP program and the test split method and compared. Third, the item difficulty and discrimination indexes (Adj Pt Bis) of the items in the data set were calculated. In the last stage, distracting analyzes of the items in the data set were made on the basis of content areas. In the IRT analysis, by using the R studio program, the data set belongs to; eta (item difficulty), thresholds threshold (the value at which the probability of a problem being made is 50%), p-value (significance value), outfit MSQ (unweighted) - infit MSQ (weighted) with outfit t (unweighted) - infit t (weighted) values were calculated respectively. In these calculations, the studies of Wu & Adams (2007) were based on in the process of testing its validity. In this context, the reverse operation of the distractors contained in the question item, the fact that the p value is less than 0.05, the MSQ value is above 1, the t value is high, and the fact that there is a low relationship between the total test score and the item leads to the classification of substances as incompatible. As a result of the analysis, it is thought that the distinctiveness and difficulty index is low, and the seven question items that do not provide the values expressed by Wu & Adams (2007) are incompatible with the model, and therefore it is considered that these question items should be reviewed.

Keywords: TIMSS, Classical Test Theory, Item Response Theory, Rasch Model, Item Analysis

³ Wu, M. & Adams, R. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach*. Educational Measurement Solutions, Melbourne.