

ÖĞRETİM MATERYALLERİNİN ÇOK YÜZEYLİ RASCH ANALİZİYLE DEĞERLENDİRİLMESİ

EVALUATION OF TEACHING MATERIALS WITH MANY-FACET RASCH ANALYSIS

Durmuş ÖZBAŞI¹, Hatice KUMANDAŞ ÖZTÜRK²

ÖZ: öğrenme araçlarının niteliği ve amacına uygunluğu ile doğrudan orantılıdır. Ayrıca öğrenmenin, birçok duyu organının uyarıcı olarak yer aldığı eğitim-öğretim ortamlarının oluşturulması eğitimdeki verimin arttıracağı göz önüne alındığında, öğretim sırasında materyal kullanımının önemini daha çok ortaya çıkarmaktadır. Bu bağlamda, araştırmanın amacını öğretmen adaylarının hazırladıkları öğretim materyallerinin belirlenen kriterler çerçevesinde farklı puanlayıcılar tarafından değerlendirilerek, puanlayıcı, kriter ve materyal yüzeylerinin Çok Yüzeysel Rasch Modeli ile incelenmesi oluşturmaktadır. Araştırmanın çalışma grubunu bir üniversitede öğrenim görmekte olan lisans öğrencileri oluşturmaktadır. Araştırmanın veri toplama aracı ise, alan yazına dayalı olarak araştırma kapsamında geliştirilen 10 maddelik materyal değerlendirme kriterleridir. Verilerin analizi için Çok Yüzeysel Rasch analizi kullanılmıştır. Araştırmanın analizlerinden elde edilen bulgulara göre; öğrencilerin rahatlıkla yapabildiği materyal kriterinin hazırlanan materyalin gerçek boyutu algısına uygun olduğu söylenebilir. Puanlayıcıların katılık/cömertlik düzeyleri açısından bazılarının çok cömert bazılarının çok sert olduğu bulunmuştur. Araştırmadan elde edilen sonuçlardan biri, öğrencilerin görsel ve metin kullanımı arasındaki denge konusunda öğrencilerin zorlandıkları, puanlayıcıların puanlamalar arasında birçok puanlama yanlışlığı yapıp, bazen bir proje katı bazen de diğer projeye cömert şekilde puanlama yaptıkları bulunmuştur.

Anahtar sözcükler: Çok yüzeysel rasch analizi, değerlendirme yanlışlığı, materyal değerlendirme, puanlayıcı güvenilirliği

ABSTRACT: The possibility of education-training processes to be carried out in a qualified manner is directly proportional to the quality and suitability of the learning tools used by the teacher. In addition, when considering that the establishment of educational environments where many sensory organs take place as stimuli will increase the efficiency in education, the importance of using materials during education is revealed even more. In this context, the aim of the research is to evaluate the teaching materials prepared by prospective teachers using different raters within the framework of determined criteria and to examine the rater, criteria and material facets using the Many-facet Rasch Model. The study group of the study consists of undergraduate students studying at a university. The data collection tool of the research is the 10-item material evaluation criteria developed within the scope of the research based on the literature. Many-facet Rasch analysis was used for the analysis of data. According to the findings obtained from the analysis of the research, it can be said that the material criterion that students can easily make is suitable for the perception of the actual size of the prepared material. It was found that some of the raters were very generous while some were very harsh in terms of severity / generosity levels. One of the results of the research was that the students had difficulties about the balance between visual and text use, that many bias were made by the raters between the scoring, and that sometimes it gave severe scoring to one project while giving a generous scoring for the other project.

Keywords: Many-facet rasch, rater bias, material evaluation, rater reliability

Bu makaleye atf vermek için:

Özbaşı, D. ve Kumandaş-Öztürk, H. (2021). Öğretim materyallerinin çok yüzeysel rasch analiziyle değerlendirilmesi, *Trakya Eğitim Dergisi*, 11(1), 187-200.

Cite this article as:

Ozbasi, D. & Kumandas-Ozturk, H. (2021). Evaluation of teaching materials with many-facet rasch analysis. *Trakya Journal of Education*, 11(1), 187-200.

¹Doç. Dr., Çanakkale Onsekiz Mart Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, dozbasi@gmail.com ORCID: 0000-0001-5078-477X.

²Doç. Dr., Artvin Çoruh Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, ORCID: 0000-0003-3636-3690.

EXTENDED ABSTRACT

Introduction

The rapid development of educational technologies in our age has caused some changes in the education process and pioneered the studies to be made for the training of qualified individuals required by the age. Using technology in the education system can create effects that facilitate learning, embody abstract concepts, provide creative and active learning, and increase motivation in individuals. One of the training programs given to prospective teachers during the teacher training process is to gain the ability to prepare appropriate learning environments and develop teaching materials. In addition, it is possible to say that the effective use of these teaching materials by the prospective teachers is important in achieving success in addition to the acquisition of this skill. In this context, in order for a prospective teacher to create an effective and active teaching environment, he/she should have the skills to prepare teaching materials that may draw the attention of the students and the skills to use educational technology, together with the ability to use it effectively in the classroom. Prospective teachers receive various training programs to improve their vocational skills during the undergraduate education process and prepare audiovisual teaching materials as required by these training programs. The teaching materials prepared can be scored by several tutors, and not just one, in line with the requirements of the course. The products in the instructional technologies and material design course, which can be performed by more than one student such as performance task, project, can be scored by multiple raters and they can be examined considering whether they have the expected criteria. Despite the numerous studies related to instructional technology and material design in Turkey in literature, there is no study examining the consistency, the reliability of the scoring done by taking into account all of the rater, measuring and scoring facets in the scoring of student products. The issue of this research is to examine the materials prepared by students within the scope of instructional technologies and material design course within the framework of material, jury and criteria. The aim of the research is to evaluate the materials prepared by prospective teachers in instructional technologies and material design course within the framework of determined criteria by different raters, and to examine the raters, criteria and material facets using the Many-faceted Rasch Model.

Within the framework of this general purpose, answers to the following questions were sought in the research.

1. In the scoring made according to the material evaluation criteria, "What is the status of the calibration map obtained for the rater, material and criteria facets?"
2. What are the statistics regarding the measurement report of the prospective teacher materials?
3. What are the statistics related to the measurement report of the criteria used in the evaluation of the materials?
4. How does the severity / generosity of the raters change during scoring for the evaluation of the materials?
5. What are the statistics about the bias analysis of the raters?

Method

It is a research in the screening model since in the study, student materials are evaluated in terms of evaluative, criterion and material dimensions and they do not aim to reveal the existing situation. In this research, 10 teaching materials prepared by prospective teachers and data obtained from 8 faculty members studying these materials were used. The materials were prepared together by groups of 8 prospective teachers. The data collection tool was created by researchers by reviewing different material review forms. In the material evaluation criteria, the suitability of the "Line, Size, Color, Texture, Integrity, Balance, Emphasis" elements and the content overlap with the achievements of the course were considered. The data analysis of the study was carried out by using the FACETS program with the Many-facet Rasch model (MFRM). SPSS 22 was also used during the control of assumptions.

Findings

When the interaction between the rater-material scores was examined, it was found that the rater number three scored the material number three as 13 points while he/she had to score 22.7 points and thus made a bias of 9.7 points. Again, while the same jury member was expected to give 40.9 points, he/she made bias by giving 50 points for the material number nine. When the bias levels of the raters are analyzed within the scope of material - jury, the same rater can behave negatively biased in the evaluation of a material while evaluating another material with a positive bias. Similarly, in the studies where the rater behaviors are examined by using MFRM, the raters are able to make positive biased scores in some individuals' and negative in others. In this study, the raters showed bias, and they scored with positive bias in some materials while demonstrating negative bias in the scoring of other materials. When the reliability coefficients obtained according to the measurement reports calculated separately for all facets in the study were examined, it was found to be between 0.80-0.97. In this regard, it can be stated that the differences between good scores in terms of reliability of the analyses performed with MFRM are reliable.

Discussion and Conclusion

According to the results obtained from the study, the material number 7 was evaluated as the most successful by the raters, while the material number 3 was considered the most unsuccessful. Another result obtained from the research is the severity / generosity of the raters regarding whether they biased during material evaluation. When the interaction between the rater-material scores was examined, it was found that the rater number three scored the material number 3 as 13 points while he/she had to score 22.7 points and thus made a bias of 9.7 points. Again, while the same jury member was expected to give 40.9 points, he/she made bias by giving 50 points for the material number nine. When the bias levels of the raters are analyzed within the scope of material - jury, the same rater can behave negatively biased in the evaluation of a material while evaluating another material with a positive bias. Similarly, in the studies where the rater behaviors are examined by using MFRM, the raters are able to make positive biased scores in some individuals' and negative in others. In this study, the raters showed bias, and they scored with positive bias in some materials while demonstrating negative bias in the scoring of other materials. When the reliability coefficients obtained according to the measurement reports calculated separately for all facets in the study were examined, it was found to be between 0.80-0.97. In this regard, it can be stated that the differences between good scores in terms of reliability of the analyzes performed with MFRM are reliable.

GİRİŞ

Çağımızda öğretim teknolojilerinin hızla gelişmesi, eğitim sürecinde de birtakım değişikliklere neden olmuş ve çağın gerektirdiği nitelikli bireylerin yetişmesi için yapılacak olan çalışmalara öncülük etmiştir. Eğitim sisteminde teknoloji kullanımı bireylerde öğrenmeyi kolaylaştırıcı, soyut kavramları somutlaştırıcı, yaratıcı ve aktif öğrenmeyi sağlayıcı, motivasyonu artırıcı etkiler oluşturabilmektedir (Rıza, 2000).

Türkiye'de uzun yıllar öğrenme ortamlarında materyal ve bilginin kaynağı olarak öğretmen ve ders kitabı kullanılmıştır. Ancak günümüzde nitelikli insan yetiştirme ihtiyacı karşısında alınan önlemler ve verilen eğitimler, bu sınırlı kaynakların kullanımını ve ihtiyacı karşılama yetersiz kılmıştır. Bu nedenle günümüzde çoklu öğrenme ortamlarına dayalı görsel, işitsel özelliklere sahip teknolojik araçların kullanımı önemli hale gelmiştir (Demirel, Seferoğlu ve Yağcı, 2004).

Geleneksel sınıf içi eğitimden farklı olarak çoklu öğrenme ortamlarının oluşturulması, kişinin sunulan bilgileri seçtiği/filtrelediği, sonrasında organize ettiği ve önceden mevcut olan bilgilerle ilişkilendirdiği aktif bir öğrenenin ortaya çıkmasını sağlamaktadır (Mayer, 2009).

Öğrenme ortamlarında görsel-işitsel öğretim materyallerinin kullanımının birçok yararının olduğu alanyazında (Apperson, Laws ve Scepanisky, 2006; Demirel, Seferoğlu ve Yağcı, 2004; Knapp ve Glenn, 1996; Rıza, 2000) üzerinde durulan önemli bir konudur. Bu görsel-işitsel öğretim materyallerinin ders kitabından farklı olarak öğrenmeyi kolaylaştırdığı, öğrenmeye karşı ilgi ve motivasyonu artırdığı, eleştirel düşünme, problem çözme ve yaratıcılık gibi üst düzey düşünme becerilerini geliştirdiği vurgulanmaktadır.

Öğretmen yetiştirme sürecinde öğretmen adaylarına verilen eğitimlerden biri de teknolojiye uygun öğrenme ortamları hazırlama ve öğretim materyali geliştirme becerisi kazandırmaktır. Ayrıca bu beceriyi kazandırma yanında öğretmen adaylarının öğretim materyallerini etkin kullanması istenen başarıya ulaşmada önemli olduğu söylenebilir. Bu bağlamda, bir öğretmen adayının etkili ve etkin bir öğretim ortamı oluşturabilmesi için, öğrencilerin ilgilerini çekebilecek eğitim materyali hazırlayabilecek ve eğitim teknolojilerini de kullanabilecek beceriye sahip olması gerektiği gibi, sınıf içerisinde de etkin bir şekilde

kullanabilmesi gerekmektedir (Varank ve Ergün, 2005). Öğretmen adayları lisans eğitimleri sürecinde mesleki becerilerini geliştirmeye yönelik çeşitli eğitimler almakta ve bu eğitimlerin gereği görsel ve işitsel öğretim materyalleri hazırlamaktadırlar. Hazırlanan öğretim materyalleri tek bir öğretici değil dersin gereklilikleri doğrultusunda birkaç öğretici tarafından puanlanabilmektedir.

Farklı puanlayıcıların yaptıkları puanlamaların güvenilirliğin belirlenmesini gerektiren pek çok ölçme durumu (performans görevi, ödevler, projeler, öğretim materyalleri, vb) mevcuttur. En az iki veya daha fazla puanlayıcının farklı bireylere ve farklı maddelere ilişkin yaptıkları puanlamalar arasındaki tutarlılığın belirlenmesi olarak tanımlanan puanlayıcı güvenilirliği (Aiken, 2000), subjektifliğin ve yanlılığın ne kadar düşük düzeyde olduğunun belirlenebilmesi için oldukça önemlidir. Birden fazla puanlayıcı ve öğretici tarafından puanlanan performans veya öğrenci ürününün ölçülmesinde puanlama tutarlılığı ve güvenilirliği bu bağlamda önemli bir husustur (Moss, 1994). Öğrencilerin hazırladıkları performans görevleri veya ürünlerinin ölçme işlemleri sırasında, birçok hata kaynağı elde edilen puanların güvenilirliğini olumsuz yönde etkileyebilmektedir (Nalbantoğlu Yılmaz ve Başusta, 2015).

Eğitimde birçok ölçme durumunun puanlanmasına subjektifliğin karışması söz konusu olabilmektedir. Bu bağlamda, farklı puanlayıcıların puanlamalarına karışan öznellik aynı zamanda bu puanlara verilen kararların doğruluğunu da etkileyebilmektedir. Bu nedenle, ölçme sonucunu gösteren puanlara dayalı verilen kararların güvenilirliğini belirlemek için, farklı puanlayıcıların verdikleri puanların birbiriyle ne derece tutarlı olduğunun belirlenmesi gereklidir (Atılğan, Kan ve Doğan, 2011). Ayrıca puanlayıcılar, puanlama anahtarlarına göre farklı şekillerde puanlamalarını yaparlar. Ancak puanlayıcıların kararları farklı faktörler/yüzeylerden etkilenmektedir. Bunlar çalışmanın kolaylık veya zorluk derecesi, puanlayıcı katılığı/cömertliği, puanlananın geçmiş vb'dir. Genellikle ölçülmesi beklenen davranışlar dışında kalan bu faktörler, puanlamanın geçerliğini ve güvenilirliğini tehdit etmektedir (Prieto ve Nieto, 2014).

Puanlayıcı güvenilirliği Klasik Test Kuramı (KTK) ve Madde Tepki Kuramına (MTK) dayalı olarak hesaplanabilmektedir. KTK'ya göre, puanlayıcı güvenilirliği puanlayıcı içi ve puanlayıcılar arası güvenilirlik kapsamında hesaplanabilmektedir. Puanlayıcı içi güvenilirlik, aynı bireyin verdiği puanların birbiriyle tutarlılığı ile ilgilidir. Puanlayıcılar arası güvenilirlik ise, birden fazla puanlayıcı arasındaki uyum veya tutarlılık olarak ifade edilebilir (Cohen ve Diğerleri, 1996). Farklı puanlayıcılar tarafından puanlanan çalışmalarla ilgili alınan kararlarda geçerlik ve güvenilirlik sorunlarına çözüm olarak KTK dışında madde tepki kuramına (MTK) dayalı yöntemler daha etkili sonuçlar vermektedir. MTK'ya dayalı yöntemlerden biri olan ve madde güçlük düzeylerini temel alan Rasch modeli, bu kuramın özel bir formu olarak bilinir (Chapman, Letourneau ve Sheidow, 2013). Rasch modellerinden biri de farklı yüzeyleri analize dahil ederek geliştirmiş olan Çok Yüzeysel Rasch Modelidir (ÇYRM, Many-facet Rasch Model-MFRM). Farklı yüzeyleri (puanlayıcı, puanlama kriteri ve puanlanan yüzeyi) analize dahil eden ÇYRM modelinin en önemli getirisi diğer yöntemlerin çözemediği ölçme sorunlarına çözüm getirmesidir (Schumacker, 1996). ÇYRM, MTK alt yapısını kullanarak, KTK'ya göre daha objektif sonuçlar elde edilmesine olanak sağlayabilmekte ve ölçme süreci içerisinde olan tüm yüzeyleri (puanlayıcı, materyal ve kriterler) dikkate alarak güvenilirlik kestirimlerinin yapılmasına olanak sağlayabilmektedir.

Performans görevi, proje gibi birden fazla öğrenci tarafından gerçekleştirilebilen Öğretim teknolojileri ve materyal tasarımı dersindeki ürünler de birden fazla puanlayıcı tarafından puanlanabilmekte ve materyalin sahip olması beklenen kriterlere sahip olup olmadığı incelenebilmektedir. Türkiye'de öğretim teknolojileri ve materyal tasarımı ile ilgili alan yazında birçok araştırma (Alım, 2013; Acer, 2011; BakaçveÖzen, 2016; Birinci, 2008; Birişçi, Metin, KaleliyilmazveCoşkun, 2011; Çalışoğlu, 2015; Çobanveİleri, 2013; GüneşveAydoğdu-İskenderoğlu, 2014; Güven, 2006; ÖzerveTunca, 2014; Yazar, 2015; Yanpar, 2009) yapılmasına karşın, öğrenci ürünlerinin puanlamasındaki puanlayıcı, ölçtü ve puanlama yüzeylerinin tümünü dikkate alarak, yapılan puanlamanın güvenilirliğini tutarlılığını inceleyen bir çalışmaya rastlanmamıştır. Bu bağlamda, bu çalışmada da öğretmen adaylarının öğretim teknolojileri ve materyal tasarımı dersinde hazırladıkları materyallerin belirlenen kriterler çerçevesinde farklı puanlayıcılar tarafından değerlendirilerek, puanlayıcı, kriter ve materyal yüzeylerinin Çok Yüzeysel Rasch Modeli ile incelenmesi amaçlanmaktadır. Bu genel amaç çerçevesinde çalışmada aşağıdaki sorulara yanıt aranmıştır.

1. Materyal değerlendirme kriterlerine göre yapılan puanlamada puanlayıcı, materyal ve kriter yüzeyleri için elde edilen kalibrasyon haritasının durumu nasıldır?
2. Öğretmen adayı materyallerine ait ölçüm raporuna ilişkin istatistikler nedir?
3. Materyallerin değerlendirilmesinde kullanılan kriterlere ait ölçüm raporuna ilişkin istatistikler nedir?

4. Materyallerin değerlendirilmesi aşamasında puanlayıcıların puanlama esnasında katılık/cömertleri nasıl değişmektedir?
5. Puanlayıcıların yanlılık analizine ait istatistikler nedir?

YÖNTEM

Araştırmanın Modeli

Araştırmada öğrenci materyallerinin değerlendirici, kriter ve materyal boyutunda ele alındığından ve var olan durumun ortaya çıkarılması amaçlanmadığından tarama modelinde bir araştırmadır. Tarama modelindeki araştırmalarda, araştırmaya konu olan değişken ve değişkenler bakımından o anki durum ortaya koyulmaya çalışılır (Karasar, 2016).

Çalışma Grubu

Bu araştırmada, farklı bölümlerde öğrenim görmekte olan öğretmen adayları tarafından hazırlanan toplam 10 öğretim materyali ve bu materyalleri inceleyen 9 öğretim üyesinden elde edilen veriler kullanılmıştır. Materyalleri 8'er kişilik öğretmen adayları grubu beraber hazırlamıştır. Bu gruplardan 2'si sosyal bilgiler, 2'si fen bilgisi öğretmenliği, 2'si Türkçe öğretmenliği, 2'si matematik öğretmenliği ve 2'si sınıf öğretmenliği bölümlerinde öğrenim gören öğrencilerden oluşmuş olup toplam 10 materyal üretilmiştir. Materyallerin konu ve içerikleri de bu bölümlere uygun olarak geliştirilmiştir. Bu materyallerin içerikleri şu şekildedir:

- Fen bilgisi öğretmenliği: Büyük ve küçük kan dolaşımı sisteminin modellenmesi, dolaşım bozukluğu oluşan doku modeli,
- Türkçe öğretmenliği: Yazar, eser ve dönem özelliği ile ilgili araç ve bilgi notu eşleştirilmesinden oluşan üç boyutlu diyagram, edebiyat ağacı (dönem, yazar ve eser eşleştirilmesi)
- Matematik öğretmenliği: Gelir ve gider hesaplamalarının somutlaştırılarak gösterildiği model, kar ve zarar hesaplama çarkı
- Sosyal Bilgiler Öğretmenliği: Bilgi türleri ve temsilcileri çarkı, televizyon makinası
- Sınıf Öğretmenliği: Maddenin hal değişimi, hücre modeli.

Bu materyaller belirlenirken amaçsal örnekleme yöntemlerinden maksimum çeşitlilik örnekleme yolu (Özbaşı, 2019) kullanılmıştır. Değerlendirici olan tüm öğretim üyeleri bu alanlarda öğretim teknolojileri ve materyal tasarımı dersini veren uzmanlardan seçilmiştir. Araştırmada projeksiyon kriteri yüzeyine uygun olarak toplam 90 (9 juri ve 10 proje matrisine göre) veri elde edilmiştir.

Veri Toplama Aracı

Veri toplama aracı farklı materyal inceleme formları gözden geçirilerek araştırmacılar tarafından oluşturulmuştur. Materyaller 10 kritere göre 5'li (Çok iyi, İyi, Orta, Kabul edilebilir ve Geliştirilmeli) olarak puanlanmaktadır. Materyal değerlendirme kriterlerinde "Çizgi, Boyut, Renk, Doku, Bütünlük, Denge, Vurgu" öğelerinin uygunluğu ve içeriğin dersin kazanımlarıyla örtüşme durumu ele alınmıştır. Bu öğeler üç boyutlu materyal geliştirirken uyulması gereken görsel tasarım öğe ve ilkelerine (Özsoy ve Ayaydın, 2015) uygun olarak seçilmiştir. Değerlendirme formu hazırlanırken alanda kullanılan formlar incelenmiş ve bu formlardan yararlanarak hazırlanan 14 madde için üç uzmandan görüş alınmış ve Davis tekniğine uygun olarak kapsam geçerliği indeksi (Davis, 1992) hesaplanmıştır. Buna göre uzmanlardan elde edilen verilerin kapsam geçerliği indeksi $KG\bar{I} = 0,80 - 1,00$ arasında olan 10 madde kapsam geçerliği için kabul edilebilir düzeyde bulunmuştur. Değerlendirme kriterleri bu 10 maddeden oluşan formdan elde edilmiştir.

Verilerin Toplanması ve Analizi

Araştırmanın veri analizinde Çok Yüzeyle Rasch modeli (ÇYRM) için FACETS programı, varsayımların kontrolü sırasında da SPSS 22 kullanılmıştır. Araştırmanın verilerini, dokuz uzman tarafından, belirlenen kriterlere göre ayrı ayrı puanlanan 10 materyalin puanlanması sonucu oluşan 90 veri oluşturmuştur.

Madde Tepki Kuramı ve Çok Yüzeyle Rasch Modeli

Verilerin analizinde Linacre (2014), tarafından geliştirilen FACETS programından yararlanılmıştır. Araştırma üç yüzeyle (projexjurixkriter) olarak 90 veri ile yürütülmüştür. Araştırmada proje yüzeyle öğrenci materyalleri, jüriyi uzman değerlendirciler ve kriterü de materyallerin değerlendirilmesinde kullanılan kriterler oluşturmaktadır. Bu bağlamda üç yüzeyle aynı anda değerlendirilebildiği veri kalibrasyon haritası ile birlikte her bir yüzeyle için ölçüm raporları hesaplanmıştır.

CYRM, madde tepki kuramının dayalı bir model olduğundan MTK'nın içerdiği varsayımları sağlaması gerekmektedir (Baker, 2001). Bu nedenle verilerin analizine başlamadan önce MTK varsayımlarını karşılayıp karşılamadığı incelenmiştir.

Tek boyutluluk (Undimensionality)

Madde tepki kuramı çok boyutlu ve tek boyutlu olmak üzere iki boyut altında incelenebilmektedir. Eğer ölçme konusu olan örtük özellikler tek boyutlu veya başat bir faktör ölçülüyorsa tek boyutlu IRT ile değilse çok boyutlu (multidimensionality IRT) modelleri (Reckase, 2009) ile incelemektedir. Bu araştırma kapsamında tek boyutlu IRT modeli kullanılmıştır. Bu bağlamda da, tek boyutluluk varsayımının sağlanıp sağlanmadığı incelenmiştir.

Bir ölçme aracının tek boyutlu olması, testteki maddeler bakımından benzer bir özelliği ölçüp ölçmediği ile ilgilidir. Testin tek boyutluluğunu belirlemenin en genel yolu, faktör analitik yöntemlerden faydalanmaktır. Faktör analitik yöntemlerden açıklayıcı faktör analizi ile çok sayıda değişken altında yatan temel örtük yapılar ortaya çıkarılabilmekte ve ölçüğün kaç boyut altında istenen örtük özelliği ölçtüğü belirlenebilmektedir (Hambleton, Swaminathan ve Rogers, 1991). AFA aynı zamanda elde edilen verilerdeki varyansın ve kovaryansın gizil kaynaklarını tespit etmek ve bunları anlamlandırmak amacıyla yapılan bir analiz tekniğidir (Jöreskop ve Sörbom, 1993). Bu araştırmada proje değerlendirme aracının tek boyutlu bir özelliğe sahip olup olmadığını anlamak amacıyla AFA yapılmıştır. Örneklem büyüklüğünün yeterliğine ilişkin KMO değeri 0.90 olarak elde edilmiş ve Barlet Küresellik testi sonucu da istatistiksel olarak anlamlı bulunmuştur ($\chi^2(45)=624,723$; $p<.01$). Buna göre verilerin faktör analizine uygun olduğu söylenebilir. Elde edilen AFA sonuçları tablo 1'de gösterilmiştir.

Tablo 1.

Proje değerlendirme aracı için AFA sonuçları

Kriter no	Faktör yükü	Kriter no	Faktör yükü	Kriter no	Faktör yükü
crt6	,86	crt10	,81	crt8	,69
crt7	,85	crt5	,80	crt3	,67
crt1	,83	crt9	,79		
crt4	,82	crt2	,74		

Öz Değer=6.204 ve Açıklanan varyans=%62

Tablo 1'de verilen AFA sonuçlarına göre, öğrenci materyallerinin değerlendirilmesinde kullanılan 10 kriter toplam varyansın %62'sini tek faktör altında açıklamaktadır. Faktör yük değerleri ise .86-.67 arasındadır. Bu bağlamda, elde edilen verilere dayalı materyal değerlendirme aracının tek boyutluluk özelliğini gösterdiği söylenebilir.

Yerel bağımsızlık (localindependence)

Maddelerin herhangi birine verilen doğru veya yanlış cevabın diğer maddenin yanıtlanmasının etkilememesi yerel bağımsızlık olarak adlandırılmaktadır (Crocker ve Algina, 1986). Bir başka ifade ile maddelerin birbirinden bağımsız bir şekilde cevaplandırılmasıdır. Hambleton ve Swaminathan (1985), yerel bağımsızlığı tek boyutluluk ile ilişkili olduğunu belirtmiş ve faktör analizi ile belirlenebileceğini belirtmiştir. Bu bağlamda, araştırmadaki verilerin faktör analizi sonucu tek boyutlu olduğu belirlendiğinden, yerel bağımsızlık varsayımını da karşılandığı ifade edilebilir.

Model veri uyumu

ÇYRM analizinde, elde edilen bulguların doğru yorumlanabilmesi için model veri uyumu varsayımının karşılanması gerekir. Model veri uyumunun karşılanabileceği için beklenmeyen değer olarak adlandırılan standartlaştırılmış artık değerler (StRes) sayısının incelenmesi gerekmektedir. Bu sayı ± 2 aralığının dışında olan StRes değerleri için tüm verinin %5'inin, ± 3 aralığının dışında kalan standartlaştırılmış artık değer sayısının tüm verinin %1'inin üzerinde olmaması gerekmektedir (Linacre, 2003a). Araştırmadaki verilerin StRes değerleri incelendiğinde bu koşulu sağladığı bulunmuş (sırasıyla bu oran %2.6 ve %0.2) ve verilerin analiz edilmesinde ve yorumlanmasında bir sorun kalmamıştır.

BULGULAR

Araştırmada elde edilen bulgular araştırmanın alt amaçlarındaki sıraya göre aşağıda verilmiştir.

Birinci alt amaca ilişkin bulgular

Araştırmanın ilk alt amacı olan, öğrenci materyallerin değerlendirilmesinde puanlayıcı, kriter ve materyal yüzeyleri için elde edilen veri kalibrasyon haritası nasıl sorusuna yanıt olarak veri kalibrasyon haritası Şekil 1'de verilmiştir.

Measr	+material	-crt	+material	-juri	-juri	MATER
2	+	+	+	+	+	(5)
			*			
7			*			
9			*			4
1	+	+	+	+	+	
			*			
		crt2				---
		crt3				
		crt5 crt8	*			
		crt9	*			
* 0 *	*	*	*	*	*	* 3 *
6		crt1 crt6	*	juri1 juri3	juri1 juri3	
5			*			
4		crt4 crt7	*			
10			*			
		crt10				---
-1	3	+	*	juri9	juri9	2
			*			
				juri4	juri4	
				juri6 juri7	juri6 juri7	
				juri5	juri5	
-2	+	+	+	juri8	juri8	---
				juri2	juri2	
-3	+	+	+	+	+	(1)
Measr	+material	-crt	* = 1	-juri	-juri	MATER

Şekil 1. Veri kalibrasyon haritası

Şekil 1’de verilen veri kalibrasyon haritasında değerlendirici, kriter ve materyaller aynı ölçek düzeyinde hesaplanarak oluşturulmuştur. Buna göre, en yüksek logit değerine sahip materyalin 7 numaralı materyal olduğu (logit değeri 1.4) söylenebilir. Kriterler incelendiğinde, öğrencilerin en yetkin oldukları kriterün de 2 numaralı kriter olan “Materyalin boyutlarının gerçek boyut algısına uygun olması” kriterü olduğu (logit değeri 0.6) ifade edilebilir. Puanlayıcılar değerlendirildiğinde ise, en katı puanlayıcının 1 ve 3 numaralı puanlayıcılar (logit değeri -0.2) olduğu, en cömert jüri üyesinin ise jüri2 kodlu puanlayıcının (logit değeri -2.5) olduğu söylenebilir. Her yüzeyi detaylı bir şekilde inceleyebilmek için her yüzeyin

ölçüm raporları hesaplanmış ve aşağıda verilmiştir. Buna göre, öğrencilerin hazırladıkları materyallere ilişkin ölçüm raporu tablo 2’de verilmiştir.

Tablo 2.

Materyal ölçüm raporu

Materyaller	Gözlenen Ort.	Yansız Ortalama	Model		Uygunluk içi		Uygunluk Dışı	
			Ölçüm	Hata	Kareler ort.	Z	Kareler ort.	Z
m7	4,71	4,78	1,41	0,21	0,75	-1,1	0,66	-1,3
m9	4,62	4,71	1,09	0,19	1,28	1,3	1,31	1,2
m2	4,48	4,58	0,69	0,16	0,96	-0,1	0,99	0
m1	4,27	4,38	0,24	0,15	1,35	1,9	1,42	2,1
m6	4,08	4,19	-0,09	0,13	1,26	1,5	1,2	1,2
m5	3,99	4,09	-0,23	0,13	0,86	-0,9	0,88	-0,7
m4	3,91	4,01	-0,35	0,13	0,63	-2,7	0,64	-2,6
m10	3,89	3,98	-0,38	0,13	0,87	-0,8	0,91	-0,5
m3	3,43	3,47	-0,99	0,12	1,13	0,9	1,16	1,1
m8	3,1	3,09	-1,39	0,11	0,84	-1,2	0,82	-1,3
Ortalama	4,05	4,13	,00	,15	,99	-0,1	1,00	-0,1
Std. Sapma	,51	,53	,88	,03	,25	1,5	,27	1,5

Model, Örneklem: RMSE .15 (Düzeltilmiş) s.d. .86 ayırıcılık 5.80 Güvenirlik .97
Model, Sabit ki-kare: 298.1 s.d.: 9 p: .000

Öğrenci materyallerine ait ölçüm raporu bulgularına göre, logit değerlerine ilişkin RMSE değeri 0.15 olarak hesaplanmıştır. RMSE değerine ait standart sapma değeri de (S.D.: 0.86) kritik değer olarak kabul edilen 1.00’in altındadır. Güvenirlik indeksi ise 0.97 olarak hesaplanmıştır. Bu bağlamda, materyal puanları arasındaki farklılığın güvenilir olduğu söylenebilir. Materyaller arasındaki farklılığın manidar olup olmadığını incelemek amacıyla ki-kare testi sonuçları incelenmiş ve materyaller arasındaki farklılığın manidar olduğu bulunmuştur ($\chi^2=298.1$, $sd=9$, $p<.05$). Bu bulguya göre, öğrenci materyalleri arasında anlamlı farklılık vardır” hipotezinde yokluk hipotezi reddedilmiştir. Bu durumda, uzmanların yapmış oldukları puanlamalara göre öğrencilerin materyallerinden elde edilen toplam puanlar arasında istatistiksel olarak anlamlı bir farkın olduğu belirtilebilir. Bir başka ifadeyle, materyaller arasında, uzmanların puanlamalarına göre nitelik açısından fark olduğu da söylenebilir. Yüzeyler ile ilgili uygunluk içi ve uygunluk dışı istatistiksel değerler incelendiğinde ise, 10 materyalin tümünün, referans aralığı olan 0.6 ile 1.4 (Wrihgt ve Lince, 1994:375-380) arasında olduğu ve bu değerleri aşan proje yüzeyinin olmadığı söylenebilir. Materyalleri değerlendiren puanlayıcılara ilişkin ölçüm raporu Tablo 3’te verilmiştir.

Tablo 3.

Puanlayıcı yüzeyine ait ölçüm raporu

Puanlayıcılar	Gözlenen Ort.	Yansız Ortalama	Model		Uygunluk içi		Uygunluk Dışı	
			Ölçüm	Hata	Kareler ort.	Z	Kareler ort.	Z
juri3	3,15	3,18	-0,19	0,11	0,89	-0,8	0,87	-0,9
juri1	3,17	3,2	-0,21	0,11	0,77	-1,9	0,8	-1,5
juri9	3,88	4	-1,12	0,12	0,64	-2,9	0,69	-2,2

juri4	4,17	4,29	-1,57	0,13	1,34	2	1,37	2
juri6	4,31	4,43	-1,83	0,14	1,08	2,6	1,42	2,1
juri7	4,32	4,44	-1,85	0,14	1,03	0,2	0,85	-0,8
juri5	4,36	4,48	-1,93	0,14	1,09	0,6	1,19	1
juri8	4,49	4,59	-2,22	0,16	0,96	-0,1	1,05	0,3
Juri2	4,58	4,67	-2,46	0,17	0,85	-0,7	0,74	-1,1
Ortalama	4,05	4,13	,000	0,15	0,99	-0,1	1	-0,1
Std. Sapma	0,51	0,53	0,88	0,03	0,25	1,5	0,27	1,5

Model, Evren: RMSE .15 (Düzeltilmiş) S.D. .82 Ayırıcılık 5.50 Güvenirlik .97
Model, Örneklem: RMSE .15 Düzeltilmiş) S.D. .86 Ayırıcılık 5.80 Güvenirlik .97
Model, Sabit ki-kare: 298.1 s.d.: 9 p: .000

Puanlayıcı yüzeyi analizlerine göre, dokuz puanlayıcıdan altı numaralı dışında diğerlerinin “uygunluk içi” ve “uygunluk dışı” değerleri incelendiğinde, Wright ve Linacre (1994:375-380) tarafından önerilen 0.6 ile 1.4 değerleri arasındadır. Öğrencilerin materyallerini değerlendiren puanlayıcılara ait ölçüm raporu bulgularına göre, RMSE değeri 0.15 ve standart sapma değeri ise örneklem için 0.86 ve popülasyon için 0.82 olarak hesaplanmıştır. Bu değerler kritik değer olarak kabul edilen 1.00’den daha düşük bir değerdir. Ayırıcılık katsayısı 8.07 olarak hesaplanmış ve Puanlayıcıların vermiş oldukları puanlar arsındaki farklılığın güvenilirliğini gösteren, güvenilirlik katsayısı ise 0.97 olarak hesaplanmıştır. ayrıca, materyal puanları arasındaki farklılığın manidar olup olmadığını incelemek amacıyla ki-kare sonuçları değerlendirildiğinde, sabit etkiye dayalı ki-kare değerinin manidar olduğu bulunmuştur ($\chi^2=298.1$, $sd=9$, $p<.01$). Buna göre, sabit etkiye ait “puanlayıcıların materyalleri değerlendirmeleri arasında anlamlı farklılık vardır” hipotezinde yokluk hipotezi reddedilmiştir. Bu bağlamda, Puanlayıcıların öğrenci materyallerini değerlendirmeleri arasında istatistiksel olarak anlamlı bir farkın olduğu sonucuna ulaşılabilir.

Materyal değerlendirme kullanılan kriterlerin analizi

Puanlayıcıların, öğrenci materyallerini değerlendirmek için kullandıkları kriterlerin amaca uygunluğunu ile ilgili çok yüzeyli rasch analizine ilişkin kriter ölçüm raporu tablo 4’te detaylı bir şekilde verilmiştir.

Tablo 4.

Materyal değerlendirme kriterlerine ilişkin ölçüm raporu

Kriterler	Gözlenen Ort.	Yansız Ortalama	Model		Uygunluk içi		Uygunluk Dışı	
			Ölçüm	Hata	Kareler ort.	Z	Kareler ort.	Z
crt8	3,76	3,91	0,48	0,13	1,09	0,6	1,14	0,8
crt5	3,83	4	0,36	0,13	1,21	1,3	1,24	1,4
crt4	3,91	4,08	0,24	0,13	1,06	0,4	1,27	1,5
crt2	3,93	4,11	0,21	0,13	1,18	1,1	1,03	0,2
crt7	4	4,18	0,1	0,14	0,97	-0,1	0,91	-0,4
crt6	4,12	4,31	-0,11	0,14	0,81	-1,1	0,92	-0,4
crt1	4,13	4,32	-0,13	0,14	1,06	0,3	1,08	0,4
crt10	4,21	4,4	-0,28	0,15	0,67	-2,1	0,75	-1,3
crt9	4,22	4,41	-0,3	0,15	0,92	-0,4	0,97	0
crt3	4,36	4,53	-0,57	0,16	0,65	-2,1	0,7	-1,4

Ortalama	4,05	4,22	0	0,14	0,96	-0,2	1	0,1
Std. Sapma	0,19	0,2	0,33	0,01	0,2	1,3	0,19	1
<i>Model, Evren: RMSE .14 (düzeltilmiş) S.D. .28 Ayırıcılık 2.03 Güvenirlik .80</i>								
<i>Model, Örneklem: RMSE .14 (düzeltilmiş) S.D. .30 Ayırıcılık 2.16 Güvenirlik .82</i>								
<i>Model, Sabit ki-kare: 49.9 d.f.: 9 p: .000</i>								

Öğrencilerin materyallerinin değerlendirilmesinde kullanılan kriterlere ilişkin ölçüm raporu tablo 4'te verilmiştir. Bu tablodaki bilgilere göre, öğrencilerin materyal hazırlama konusunda en çok zorlandıkları kriterlerin başında 8 numaralı kriter olan “Materyalde görsel ve metinler için denge unsuruna yer verilmesi” gelmektedir. Öğrencilerin materyal hazırlarken en az zorlandıkları veya en fazla puan aldıkları kriter ise üç numaralı kriter olan “Materyaldeki dokunun görsel algılamayı kolaylaştırması”dır. Kriter yüzeyi ölçüm raporu sonuçlarına göre, ayırma indeksi 3.22, güvenirlilik katsayısı ise 0.82 olarak hesaplanmıştır. Sabit etkiye ait “materyal değerlendirmede kullanılan kriterlerin güçlükleri arasında anlamlı bir fark vardır” hipotezini test eden ki-kare sonuçlarına göre de anlamlı fark vardır ($\chi^2=49.9$, $sd=9$, $p<.05$). Buna göre, yokluk hipotez reddedilmiş olup, öğrencileri materyalleri değerlendirmede kullanılan kriterlerin güçlük/kolaylıkları açısından aralarında istatistiksel olarak anlamlı bir farkın olduğu söylenebilir.

Puanlayıcı-materyal yanlılık analizi

Öğrencilerin materyallerini değerlendiren puanlayıcıların, puanlama sırasında yanlılık yapıp yapmadıkları ya da katılık/cömertlik durumlarını inceleyebilmek için puanlayıcılar ve materyal arasındaki yanlılık analizinin incelenmesi gerekmektedir. Puanlayıcı-materyal yanlılık analizi tablo 5'te verilmiştir.

Tablo 5.

Jüri-materyal yanlılık analizi

Gözlene n Puan	Beklene n puan	Gözle m Sayısı	Gözlene n- Beklene n Ortalama	Yanlılık modeli Büyüklüğü	Standart hata	Uygunlu k içi Kereler ortalama sı	Uygunlu k dışı Kereler ortalama sı	jürile r	Matery al
33	42,61	10	-0,96	-1,23	0,33	0,7	0,7	juri5	m10
24	34,29	10	-1,03	-1,11	0,34	0,7	0,7	juri7	m8
13	22,7	10	-0,97	-1,81	0,61	0,7	0,7	juri3	m3
35	42,21	10	-0,72	-0,94	0,33	1,9	1,8	juri4	m6
35	42,2	10	-0,72	-0,94	0,33	0,2	0,2	juri6	m4
37	42,89	10	-0,59	-0,83	0,34	1,4	1,5	juri6	m5
38	43	10	-0,5	-0,73	0,35	0,3	0,3	juri1	m7
43	46,6	10	-0,36	-0,84	0,42	1,2	1,2	juri8	m1
44	37,77	10	0,62	0,94	0,45	1,8	2,2	juri7	m3
46	38,42	10	0,76	1,33	0,53	1	0,8	juri5	m3
46	37,61	10	0,84	1,43	0,53	1	0,9	juri6	m3
50	48,21	10	0,18	1,32	< 1,42	0,1	0,1	juri2	m2
50	43,66	10	0,63	2,73	< 1,42	0,1	0,1	juri6	m6
50	48,5	10	0,15	1,13	< 1,42	0,1	0,1	juri5	m7
50	48,35	10	0,17	1,23	< 1,42	0,1	0,1	juri6	m7
50	48,38	10	0,16	1,21	< 1,42	0,1	0,1	juri7	m7

50	40,9	10	0,91	3,19	< 1,42	0,1	0,1	juri3	m9
50	47,81	10	0,22	1,53	< 1,42	0,1	0,1	juri7	m9

Sabit (tümü ki-kare: 167.2 s.d.: 90 p : .000)

Tablo 5'te Puanlayıcıların öğrenci materyallerine verdikleri puanlarda yanlılık olup olmadığına ilişkin bulgulara yer verilmiştir. ÇYRM analizlerinde yanlılık için t değerlerinin incelenmesi gerekmektedir. Bu t değerinin de yanlı buna göre, araştırmaya katılan planlayıcılardan jüri3 ve jüri 6 kodlu olanların yanlı davrandığı söylenebilir. Linarce (2014)'e göre, t değerinin -2 ile +2 dışında olması puanlayıcının yanlı puanlama yaptığını göstermektedir. Bu bağlamda, 3 numaralı puanlayıcının 9 numaralı materyali puanlarken 40.9 puan vermesi gerekirken 50 puan vererek cömert bir puanlama yapmıştır. Benzer bir durum 6 numaralı puanlayıcı için de söylenebilir. Buna göre, 6 numaralı puanlayıcı, 43.66 puan vermesi gerekirken 50 puan vererek cömert bir puanlama yaptığı belirtilebilir. Diğer puanlamalar için beklenen ve gözlenen puan arasındaki fark göz ardı edilebilir düzeyde olduğu için tam olarak yanlılık yapıldığı söylenemez.

TARTIŞMA, SONUÇ ve ÖNERİLER

Bu araştırmada, öğrencilerin Materyal Geliştirme dersi kapsamında hazırlamış oldukları materyaller, çeşitli kriterlere göre ÇYRM analiziyle incelenmiştir. Araştırma bulgularından elde edilen sonuçlara göre, hazırlanan 10 materyalden 7 numaralı materyal puanlayıcılar tarafından en başarılı olarak değerlendirilirken, 3 numaralı materyal en başarısız olarak değerlendirilmiştir. Puanlayıcıların katılık ve cömertlikleri incelendiğinde ise, en cömert puanlayıcının iki, en katı puanlayıcının ise üç numaralı jüri üyesi olduğu bulunmuştur.

Materyallerin değerlendirildiği kriterler açısından elde edilen sonuçlar incelendiğinde ise, uzmanların puanlamalarına göre sekiz numaralı kriter, öğrencilere en zor gelen kriter olarak bulunmuştur. Sekiz numaralı kriter, "Materyalde görsel ve metinler için denge unsuruna yer verilmesi" kriterüdür. Bir sonuca göre, öğrenciler materyallerinde görseller ile metin arasında denge kurmakta zorlandıkları da söylenebilir. Setterhwaite 'a (1990) göre, öğretim materyalinin, öğretim hedeflerine yönelik olarak görsel ve işitsel unsurları içerecek şekilde önceden hazırlanması gerekmektedir. İyi tasarlanmış bir görsel ile sözcüklerden daha etkili ve kalıcı bir öğrenmenin gerçekleşmesi daha mümkün olacağı araştırmacılar tarafından ifade edilmektedir (Erişti, 2005; Grabinger, 1999; Jonassen, 2002). TedCobun'a göre, öğrenilenlerin % 83'ünü görme, % 11'ini işitme, % 3,5'ini koklama, % 1,5'ini dokunma, % 1'ini duyularıyla elde ettiği yaşantılar yoluyla gerçekleşmektedir (Çilenti, 1992). Ayrıca materyal içerisindeki görsel öge, öğrencinin yönlendirilmesinde, dikkatini toplamasında, analiz ve sentez yapabilmesinde yardımcı olabilmektedir (Genty'e (1994). Bu nedenle materyale uygun bir görselin yerleştirilmesi son derece önemlidir. Çünkü görselin hedefe uygun olmasının yanında metin ile de uyumlu olması gerekmektedir. Bu araştırma sonuçlarına göre de öğrenciler materyal hazırlarken materyalde görsel ve metinler arasında denge kurmada zorlandıkları bulunmuştur.

Araştırmadan elde edilen bir başka sonuç ise, puanlayıcıların materyal değerlendirme sırasında yanlılık yapıp yapmadıkları ile ilgili olarak katılık/cömertlikleridir. Puanlayıcı-materyal puanları arasındaki etkileşim incelendiğinde, üç numaralı puanlayıcının üç numaralı materyali puanlarken 22,7 puan vermesi gerekirken, 13 puan vererek 9,7 puanlık yanlılık yaptığı bulunmuştur. Yine aynı jüri üyesinin ise dokuz numaralı materyali puanlarken ise, 40,9 puan vermesi beklenirken 50 puan vererek yine yanlı davranmıştır. Puanlayıcıların materyal – jüri kapsamında genel olarak yanlılık düzeyleri incelendiğinde, aynı puanlayıcı bir materyali pozitif anlamda yanlı değerlendirirken bir başka materyalin değerlendirmesinde negatif anlamda yanlı davranabilmektedir. ÇYRM kullanarak, puanlayıcı davranışlarının incelendiği araştırmalarda (Köse, Usta ve Yandı, 2016; Baştürk ve Işıkoğlu, 2007 da benzer şekilde puanlayıcılar bazı bireylerin çalışmalarında pozitif, bazılarıninkinde ise negatif anlamda yanlı puanlama yapabilmektedirler. Bu araştırmada, puanlayıcılar yanlılık göstermiş, bazı materyallerin puanlanmasında pozitif bazılarıninkinde ise negatif yönlü yanlılık yaparak puanlama yapmışlardır. Araştırmadaki tüm yüzeylere ait ayrı ayrı hesaplanan ölçüm raporlarına göre elde edilen güvenilirlik katsayıları incelendiğinde 0.80-0.97 arasında olduğu bulunmuştur. Bu bakımdan, ÇYURM ile gerçekleştirilen analizlerin güvenilirlik açısından iyi düzeyde (Şencan, 2005), puanlamalar arasındaki farkların güvenilir olduğu belirtilebilir. Linacre (2014), ÇYRM ile hesaplanan güvenilirlik katsayısı klasik test kuramından farklı olarak, yapılan puanlamalar

arasındaki farkın güvenilirliği olarak yorumlanması gerektiğini belirtmektedir. Bu bağlamda, araştırmada elde edilen güvenilirlik sonuçlarına göre, farkların güvenilir olduğu yorumu yapılabilir.

Araştırmanın sonuçlarına dayalı olarak bundan sonra, değerlendirme yapacak olan puanlayıcılara ölçütler ve değerlendirmenin nasıl yapılacağı ile ilgili eğitim verilmesi, puanlamaların daha güvenilir ve yansız olmasını sağlayabilecektir. Ayrıca, ileride yapılacak araştırmalar için ise, puanlayıcıların verilerini ÇYRM ile incelemenin yanı sıra puanlayıcılar arası güvenilirliğin incelenmesi ve her iki yöntemle elde edilen verilerin karşılaştırılarak analiz edilmesi de önerilebilir.

KAYNAKÇA

- Aiken, L. R. (2000). *Psychological Testing and Assessment*. Boston: Allyn and Bacon.
- Acer, D. (2011). Okulöncesi öğretmen adaylarının materyal geliştirme dersine ilişkin görüşlerinin incelenmesi. *İlköğretim Online*, 10(2), 421-429.
- Alım, M. (2013). Coğrafya öğretmeni adaylarının öğretim teknolojileri ve materyal tasarımı/geliştirme dersinde elde ettikleri kazanımlar. *Doğu Coğrafya Dergisi* – 33.
- Apperson, J. M., Laws, E. L., & Scepansky, J. A. (2006). The impact of presentation graphics on students' experience in the classroom. *Computers and Education*, 47(1), 116-126.
- Atılğan, H., Kan, A., & Doğan, N. (2011). *Eğitimde ölçme ve değerlendirme*. Anı Yayıncılık
- Bakaç, E. & Özen, R. (2016). Öğretmen adaylarının öğretim teknolojileri ve materyal tasarımı dersine yönelik tutumları, yaratıcılık algıları ve öz-yeterlik inançları arasındaki ilişki. *İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 16 (1), 41-61.
- Birinci, E. (2008). *Materyal tasarımı ve geliştirilmesinde proje tabanlı öğrenmenin kullanılmasının öğretmen adaylarının eleştirel düşünme, yaratıcı düşünme ve bilimsel süreç becerilerine etkisi*. Yüksek Lisans Tezi, Zonguldak Karaelmas Üniversitesi Sosyal Bilimler Enstitüsü, Zonguldak.
- Birişçi, S., Metin, M., Kaleliyılmaz, G. & Coşkun, K. (2011). Öğretim materyallerine yönelik web sayfalarını tasarlarken öğretmen adaylarının karşılaştıkları sorunlar, *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, 17, 102-118.
- Çalışoğlu, M. (2015). Sınıf öğretmeni adaylarının öğretim teknolojileri ve materyal tasarımı dersine ilişkin görüşleri. *CurrRes Educ*, 1(1),s.23-32
- Chapman, J. E., McCart, M. R., Letourneau, E. J., & Sheidow, A. J. (2013). Comparison of youth, caregiver, therapist, trained, and treatment expert raters of therapist adherence to a substance abuse treatment protocol. *Journal of Consulting and Clinical Psychology*, 81(4), 674.
- Çilenti, K. (1992). *Eğitim teknolojisi ve öğretim*. Ankara: Kadioğlu Matbaası.
- Çoban, A. & İleri, T. (2013). Sosyal bilgiler öğretmenlerinin öğretim teknolojileri ve materyalleri kullanma düzeyleri ve kullanamama sebepleri. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 2(1), 194-213.
- Demirel, Ö., Seferoğlu, S.S., & Yağcı, E. (2004). *Öğretim teknolojileri ve Materyal Geliştirme*. Ankara: PegemA.
- Erişti, S. D. (2005). *Grafik tasarım ilkelerine dayalı olarak geliştirilmiş etkileşimli eğitim cd'lerinin ilköğretimde temel sanat elemanlarının öğretiminde etkililiği*. Doktora tezi, Anadolu Üniversitesi Eğitim Bilimleri Enstitüsü. Eskişehir.
- Genty, C. G. (1994). *Introduction to instructional development: Process and technique*. Belmont California: Wadsworth PUBLISHING Company.
- Grabinger, S. (1999). *Instructional strategies in distance science courses: Can the web improve undergraduate science education?* İnternet adresi: [http:// web. Uccs. edu/bgaddis/leadership/litreviewD2.html](http://web.Uccs.edu/bgaddis/leadership/litreviewD2.html). adresinden 03.07.2005 tarihinde alınmıştır.
- Güneş, G., & Aydoğdu İskenderoğlu, T. (2014). İlköğretim matematik öğretmeni adaylarının öğretim teknolojileri ve materyal tasarımı dersine yönelik yaklaşımları. *GEFAD /GUGJEF*, 34(3): 469-488.
- Güven, S. (2006). Öğretim teknolojileri ve materyal geliştirme dersinin kazandırdığı yeterlikler yönünden değerlendirilmesi (inönü üniversitesi eğitim fakültesi örneği). *Türk Eğitim Bilimleri dergisi*, 4 (2), 165-179. Retrieved from <http://dergipark.org.tr/tebd/issue/26120/275167>
- Karasar, N. (2016). *Bilimsel araştırma yöntemi: Kaynaklar ilkeler teknikler*. Ankara: Nobel yayıncılık.
- Knapp, L. R., & Glenn, A. D. (1996). *Restructuring schools with technology*. Boston: Allyn and Bacon.

- Linacre, J. M. (2003). Size vs. significance: Standardized chi-square fit statistic. *Rasch Measurement Transactions*, 17(1), 918.
- Linacre, J.M. (2014). *A user's guide to FACETS Rasch-model computer programs*. 30 Aralık 2016 tarihinde <http://www.winsteps.com/a/facets-manual.pdf> adresinden alınmıştır.
- Özbaşı, D. (2019). Örneklem Belirleme. K. Yılmaz & R. S. Arık (Ed.), *Eğitimde Araştırma Yöntemleri* içinde (102-126), Ankara: PegemAkademi.
- Özer, Ö. & Tunca, N. (2014). Öğretmen adaylarının materyal hazırlama ve kullanmaya yönelik görüşleri. *Route Educational and Social Science Journal*, 1(3), 214-229.
- Özsoy, V. & Ayaydın, A. (2015). *Görsel tasarım öge ve ilkeleri*. Ankara: Pegem Akademi.
- Prieto, G. & Nieto, E. (2014). Analysis of rater severity on written expression exam using many faceted *Rasch measurement*. *Psicologica*, 35, 385-397.
- Reckase, M. D. (2009). *Multidimensional Item Response Theory*. New York, NY: Springer.
- Rıza, E. T. (2000). *Eğitim teknolojisi uygulamaları ve materyal geliştirme*. İzmir: Anadolu Matbaası.
- Schumacker, R. E. (1996). *Many-Facet Rasch Model Selection Criteria: Examining Residuals and More*.
- Setterhwaite, L. (1990). *Instructional media: Materials production and utilization*. KendalHurt Pupliching Company.
- Varank, İ. & Ergün, S. S. (2005). Eğitim Teknolojisi ve Materyal Geliştirme Yeterliliklerinin Belirlenmesi ve Eğitim Fakültesi Son Sınıf Öğrencilerinin Bu Yeterlilikle Sahip Olma Derecelerinin İncelenmesi İçin Bir Araştırma Önerisi. *Ulusal Eğitim Bilimleri Kongresi 14. Sempozyumu*, (1), 838-840. Pamukkale Üniversitesi Eğitim Fakültesi, Denizli.
- Yazar, T. (2015). Öğretmen adaylarının öğretim teknolojileri ve materyal tasarımı dersi hakkındaki görüşleri. *Uluslararası Eğitim Programları ve Öğretim Çalışmaları Dergisi*, 5(9).
- Yanpar, T. (2009). Öğretmen adaylarının portfolyoları üzerinde grup olarak yaratıcılık temelli materyal geliştirmenin etkileri. *Eğitim ve Bilim*, 34(153), 83-98.