

PAPER DETAILS

TITLE: Investigation of Achievement Test Development Competencies of Primary Mathematics Teacher Candidates

AUTHORS: Mehmet SATA

PAGES: 42-56

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/2484582>

Investigation of Achievement Test Development Competencies of Primary Mathematics Teacher Candidates

Mehmet ŞATA, Agri Ibrahim Cecen University, ORCID ID: 0000-0003-2683-4997

Abstract

This study aimed to examine the achievement test development competencies in the sample of primary mathematics teacher candidates since the competencies of teacher candidates regarding in-class assessment and evaluation activities have a crucial role in the education system. For this purpose, the descriptive model, one of the quantitative research approaches, was used. The research was carried out with 57 primary mathematics teacher candidates. Twenty-four groups were formed from the teacher candidates within the scope of the research, and each group developed an achievement test. The research data were obtained using the analytical rubric for the achievement test developed by the researcher. The developed analytical rubric has six criteria and a five-point rating. Exploratory factor analysis, Cronbach's alpha coefficient, and the many-facet Rasch model were used in data analysis. When the study's findings were examined, it was found that the teacher candidates who developed the achievement test had different proficiency levels. In addition, it was determined that the most challenging stage in the development of the achievement test was writing the appropriate item for the learning outcome, while the least difficult stage was determining the purpose of the achievement test. It was determined that the groups with low proficiency mainly chose the acquisitions for high-level mental skills and therefore had difficulty writing questions. The findings were discussed with the results in the literature. In order to support teacher candidates' question writing competencies, it is necessary to increase the application dimension by giving elective courses for measurement and evaluation applications.

Keywords: Teacher competencies, measurement, and assessment, achievement test, many facet Rasch



Inonu University
Journal of the Faculty of
Education
Vol 24, No 1, 2023
pp. 42-56
[DOI](#)
10.17679/inuefd.1130273

[Article Type](#)
Research Article

[Received](#)
13.06.2022

[Accepted](#)
27.12.2022

Suggested Citation

Şata, M. (2023). Investigation of achievement test development competencies of primary mathematics teacher candidates, *Inonu University Journal of the Faculty of Education*, 24(1), 42-56. DOI: 10.17679/inuefd.1130273

This article was presented as an oral presentation at the Measurement and Assessment in Education National Congress on Applications held in İstanbul on 29-30 May 2021.

EXTENDED ABSTRACT

Introduction

There should be objectives, content, learning-teaching process, and assessment and evaluation elements in a well-designed curriculum. In this context, there is a need for well-designed in-class assessment and evaluation practices for achieving the goals set by the program and for the behavior change of students (Karaduman & Yelken, 2020). Undoubtedly, the person who carries out the in-class assessment and evaluation activities is the teacher. Because the classroom measurement and evaluation activity most appropriate to the student's cognitive level is carried out only by the teacher responsible for the classroom and who knows the classroom, the teacher should have the qualifications in the field of measurement and evaluation in order to use classroom measurement and evaluation activities effectively (Turgut & Baykul, 2010).

When the literature is examined, it has been determined that the proficiency levels of teacher candidates for measurement and assessment activities are low (Altun & Gülay, 2017). The basis of this situation is the theoretical weight of undergraduate education (Boyras, 2007) and adaptation problems (Korkmaz et al., 2004). Besides, it was determined that candidate teachers' proficiency in some classroom measurement and assessment tools was lower than in other assessment tools (İnceçam et al., 2018). In this context, in the related study, multiple-choice item formats were chosen in the achievement tests prepared by the candidate teachers. The national wide-scale exams are in multiple-choice item format, and the teachers' scoring skills are low.

Purpose

Although the importance of teacher candidates' achievement test development competencies and their impact on student achievement, it is essential to determine these competencies. In this context, the primary purpose of the related study has been determined as the examination of candidate teachers' achievement test development competencies.

Method

This study aimed to evaluate the achievement test development competencies of teacher candidates. For this purpose, the survey model, one of the quantitative research approaches, was used (Karasar, 2012). The study group consists of 57 teacher candidates in the primary mathematics department of the education faculty of a university located in the eastern Anatolia region. In the study, teacher candidates formed groups of two and three (a total of 24 groups), and achievement tests were developed by these groups. The achievement tests developed by the teacher candidates were scored three times by the researcher. The scoring process was carried out over two months. Only intra-rater reliability was calculated in the study because a single rater made evaluations at different times. The data within the scope of research were collected with an analytical rubric to evaluate the competence of developing an achievement test. The analytical rubric consists of 6 criteria (purpose of the test, using the specification table, writing the questions, expert opinion, item analysis, and item selection, reporting, and final form) and five categories (1 = very inadequate, 2 = inadequate, 3 = moderate, 4 = adequate, 5 = excellent) (see Appendix 1). Many-facet Rasch measurement model was used

in data analysis. Unidimensionality, local independence, and model data fit, which are the assumptions of the many-facet Rasch measurement model, were tested, and it was determined that the assumptions were met. Analysis were carried out under a fully crossed design. As a result of the exploratory factor analysis, it was found that the measurement tool was one-dimensional.

Findings

In this study, which aimed to examine the achievement test development competencies of teacher candidates, the competency levels of the teacher candidates who developed the achievement test were examined. When the logit and observed average values in analysis are examined, it is seen that the group with the highest level of proficiency is number 15, and the group with the lowest level of competence is number four. It has been found that the reliability of the established model's separation rate, separation index, and separation index is high. These values are high, indicating that the pre-service teachers who developed the achievement test have different proficiency levels. In addition, the fact that the fixed effect chi-square value is significant indicates that this difference is statistically significant. Since logit values do not have any threshold value, t-values are obtained by using logit values. Obtained t-values were compared with critical t-values and significant results were determined ($t_{critical} (df) = 2.064 (23); p = .05$).

After examining the proficiency level of each group that developed the achievement test, the analytical rubric criteria were examined. It was tried to determine which of the criteria the teacher candidates were better or worse. It is seen that the criterion that teacher candidates have the lowest level of proficiency or the most difficulty in the process of developing an achievement test is clear question writing (logit = -0.43). On the other hand, it is seen that the criterion with which the pre-service teachers have the most serious or slightest difficulty in the process of developing an achievement test is determining the purpose of the achievement test (logit = 0.84).

Discussion & Conclusion

Within the scope of the research, the proficiency levels of the 24 achievement tests prepared by 57 primary mathematics teacher candidates were examined. It was determined that five groups within 24 groups were low proficiency and three groups were high proficiency. When the literature is examined, it is seen that teachers and teacher candidates have low levels of proficiency in assessment and evaluation activities (Gaitas & Alves-Martins, 2017; Şenel et al., 2018). In addition, teachers and teacher candidates avoid preparing an achievement test because they think it will take much time to prepare, apply and determine the situation (Çakan, 2004; Topkaya & Yılar, 2016). The fact that national large-scale exams consist of multiple-choice items causes teachers and teacher candidates to tend only to this type of question (Bayat & Şentürk, 2015). In this context, it is seen that results obtained within the scope of the current research are supported in the literature.

In addition, it has been determined that the achievements of the groups with low proficiency levels regarding the achievement tests have the achievements of high-level cognitive thinking skills. It is stated in the literature that it is challenging to write multiple-choice items for high-level cognitive skills (Kutlu et al., 2014; Moore, 2009). In this context, it is thought that the teacher candidates' competencies in preparing for achievement tests are lower than they should

be due to this reason. From this point of view, the development and evaluation of the achievement test for different achievements are considered the study's limitations. In addition, it is thought that teacher candidates have difficulties in transforming the acquisitions of high-level mental skills into a question format since they have not written multiple-choice items and developed an achievement test before.

Based on these results, it is thought that, unlike the current program of the measurement and evaluation course taught in education faculties, the weekly course hours should be at least doubled, and the application dimension should be as much as the theoretical dimension. In addition, it is expected that the question writing training given to the teacher candidates will contribute to the more valid and reliable measurements.

İlköğretim Matematik Öğretmen Adaylarının Başarı Testi Geliştirme Yeterliliklerinin İncelenmesi

Mehmet ŞATA, Ağrı İbrahim Çeçen Üniversitesi, ORCID ID: 0000-0003-2683-4997

Öz

Öğretmen adaylarının sınıf içi ölçme ve değerlendirme faaliyetlerine ilişkin yeterliliklerin eğitim sisteminde kilit role sahip olmasından dolayı, bu çalışmada ilköğretim matematik öğretmen adayları örnekleminde başarı testi geliştirme yeterliliklerinin incelenmesi amaçlanmıştır. Bu amaç doğrultusunda nicel araştırma yaklaşımlarından biri olan betimsel modelden faydalanılmıştır. Araştırma 57 ilköğretim matematik öğretmen adayı ile gerçekleştirilmiştir. Araştırma kapsamındaki öğretmen adaylarından 24 grup oluşturulmuş ve her grup bir başarı testi geliştirmiştir. Araştırma verisi, araştırmacı tarafından geliştirilmiş olan başarı testi için analitik dereceli puanlama anahtarı kullanılarak elde edilmiştir. Geliştirilen analitik dereceli puanlama anahtarı altı ölçüt ve beşli derecelendirmeye sahiptir. Veri analizinde açıklayıcı faktör analizi, Cronbach alfa katsayısı ve çok yüzeysel Rasch modeli kullanılmıştır. Araştırmanın bulguları incelendiğinde, başarı testi geliştiren öğretmen adaylarının farklı yeterlilik düzeylerine sahip oldukları bulunmuştur. Ayrıca başarı testinin geliştirilmesi sürecinde en fazla zorlandıkları aşamanın kazanıma uygun madde yazımı iken, en az zorlandıkları aşama ise başarı testinin amacının belirlenmesi olduğu tespit edilmiştir. Yeterliliği düşük olan grupların daha çok üst düzey zihinsel becerilere yönelik kazanımları seçtikleri ve bundan dolayı soru yazmada zorlandıkları belirlenmiştir. Elde edilen bulgular alanyazındaki sonuçlar ile tartışılmıştır. Öğretmen adaylarının soru yazma yeterliklerinin desteklenmesi amacıyla ölçme ve değerlendirme uygulamalarına yönelik seçmeli derslerin verilerek uygulama boyutunun artırılması gerekmektedir.

Anahtar Kelimeler: Öğretmen yeterlikleri, ölçme ve değerlendirme, başarı testi, many facet Rasch



İnönü Üniversitesi
Eğitim Fakültesi Dergisi
Cilt 24, Sayı 1, 2023
ss. 42-56

DOI
10.17679/inuefd.1130273

Makale Türü
Araştırma Makalesi

Gönderim Tarihi
13.06.2022

Kabul Tarihi
27.12.2022

Önerilen Atıf

Şata, M. (2023). İlköğretim matematik öğretmen adaylarının başarı testi geliştirme yeterliliklerinin incelenmesi. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, 24(1), 42-56. DOI: 10.17679/inuefd.1130273

Bu makale 29-30 Mayıs 2021 tarihlerinde İstanbul'da gerçekleştirilen Eğitimde Ölçme Değerlendirme Uygulamaları Ulusal Kongresi'nde özet bildirisi olarak sunulmuştur.

İlköğretim Matematik Öğretmen Adaylarının Başarı Testi Geliştirme Yeterliliklerinin İncelenmesi

İyi tasarlanmış bir eğitim programında, hedefler, içerik, öğrenme-öğretme süreci ve ölçme değerlendirme unsurları bulunmalıdır. Bu bağlamda programın belirlemiş olduğu hedeflere ulaşma, öğrencilerin davranış değişikliği gerçekleşme durumları için iyi tasarlanmış sınıf içi ölçme ve değerlendirme uygulamalarına ihtiyaç duyulmaktadır (Karaduman ve Yelken, 2020). Sınıf içi ölçme ve değerlendirme faaliyetini en iyi yürüten kişi şüphesiz ki öğretmen olmaktadır. Çünkü öğrencinin bilişsel seviyesine en uygun sınıf içi ölçme ve değerlendirme faaliyeti, ancak sınıftan sorumlu ve sınıfı bilen öğretmen tarafından gerçekleştirilmektedir. Öğretmenin sınıf içi ölçme ve değerlendirme faaliyetlerini etkin bir biçimde kullanabilmesi için ölçme ve değerlendirme alanındaki yeterliliklere sahip olması beklenmektedir (Turgut ve Baykul, 2010).

Eğitim programının hedeflerine ulaşması, öğrenci başarısının belirlenmesi, öğrenci öğrenmelerinin zayıf ve güçlü yönlerinin belirlenmesi ve eğitim-öğretim sürecinin etkililiği gibi eğitim sisteminin girdi ve çıktısında sınıf içi ölçme ve değerlendirme faaliyetlerinden faydalanılmaktadır (Güler ve Gelbal, 2010). Bu bağlamda öğretmenlerin ölçme ve değerlendirme faaliyetlerinin değerlendirilmesi önem arz etmektedir. Çünkü öğretmenin sahip olduğu ölçme ve değerlendirme yeterliliği öğrencilerin öğrenme durumlarını doğrudan etkilemektedir (Boud, 1995). Aynı zamanda eğitim programında belirlenen hedef ve kazanımlara ulaşmada da ölçme ve değerlendirme uygulamalarının kullanım düzeyi belirleyici bir role sahiptir (Smith ve Southerland, 2007).

Ülkemizde ölçme ve değerlendirme faaliyetlerinde temel amaç öğrenci veya öğrenenin başarısını belirlemektir. Bu bağlamda kullanılan yöntemler geleneksel ve tamamlayıcı ölçme ve değerlendirme olarak iki genel sınıfa ayrılmaktadır. 2005 yılından beri yapılandırmacı yaklaşıma geçiş yapılmasına rağmen eğitim sistemimizde ve ulusal geniş ölçekli başarı belirleme sınavlarında geleneksel yaklaşımın daha yaygın olduğu görülmektedir. Hem sınıf içi hem de geniş ölçekli başarı düzeylerinin belirlenmesi sürecinde geleneksel yöntemlerden çoktan seçmeli başarı testlerinin kullanıldığı görülmektedir (Açıkgöz ve Karslı, 2015; Akbulut ve Çepni, 2013). Bu bağlamda eğitim fakültesinden mezun olan öğretmen adaylarının başarı testi geliştirme süreçlerinin durum belirlenmesinin yapılması önemli görülmektedir.

İyi bir başarı testinin sahip olması gereken nitelikleri bulunmaktadır. Özellikle ölçümlerin güvenilir ve geçerli olması temel bir niteliktir. Başarı testlerinin geliştirilmesi sürecinde sistematik bir sürecin takip edilmesi elde edilen ölçümlerin güvenilirlik ve geçerliğine katkı sağlayacaktır. Bir başarı testinde hazırlık, uygulama ve analiz olmak üzere üç temel aşama bulunmaktadır. Hazırlık aşamasında ilk olarak başarı testinin amacı belirlenir. Daha sonra bu amaç doğrultusunda ölçülecek davranış evreni belirlenmektedir. Davranış evrenini temsil eden davranış örnekleme seçildikten sonra madde yazım sürecine geçilmektedir. Madde havuzunun oluşturulması süreciyle birlikte hazırlık aşaması tamamlanıp uygulama aşamasına geçilmektedir. Bu aşamada ilk olarak uzman görüşleri alınarak ölçümlerin kapsam geçerliğine kanıt sağlanmaktadır. Uzman görüşü doğrultusunda revize edilen ölçme aracı küçük bir grup ile ön pilot çalışması yapılmaktadır. Böylelikle hazırlanan soruların anlaşılabilirliği ve ne kadar zaman içinde çözüleceği gibi teknik sorunlara yanıtlar bulunur. Bu adım tamamlandıktan sonra ana uygulamaya geçilir. Soruların yazıldığı grubun belli bir kısmına sorular uygulanır. Son olarak analiz ve raporlama sürecinde her bir madde için analizler gerçekleştirilip uygun olanlar ölçme aracında dahil edilir.

son hali verilir. Bu araştırmada da ölçme aracının geliştirilme süreci öğretmen adayları ile paylaşılmış ve tüm aşamaların yapılması sağlanmıştır. Böylece elde edilen ölçümlerin daha güvenilir ve geçerli olması sağlanmıştır.

Alanyazın incelendiğinde, öğretmen adaylarının ölçme ve değerlendirme faaliyetlerine yönelik yeterlilik düzeylerinin düşük olduğu tespit edilmiştir (Altun ve Gülay, 2017). Bu durumun temelinde, lisans eğitiminin teorik ağırlıklı olması (Boyraz, 2007) ve uyum sağlama problemleri (Korkmaz et al., 2004) olduğu görülmektedir. Ayrıca öğretmen adaylarının bazı sınıf içi ölçme ve değerlendirme araçlarındaki yeterliliklerinin diğer ölçme araçlarına göre daha düşük olduğu belirlenmiştir (İnceçam ve ark., 2018). Bu bağlamda ilgili çalışmada öğretmen adaylarının başarı testi geliştirme sürecinde, hem ulusal çaptaki geniş ölçekli sınavların çoktan seçmeli madde formatında olması hem de öğretmenlerin puanlama becerilerinin düşük olmasından dolayı, öğretmen adayları tarafından hazırlanan başarı testlerinde çoktan seçmeli madde formatları seçilmiştir.

Öğretmen adaylarının başarı testi geliştirme yeterliliklerinin önemi ve öğrenci başarı belirleme üzerindeki etkisi dikkate alındığında bu yeterliliklerin belirlenmesi önemli görülmektedir. Bu bağlamda ilgili çalışmanın temel amacı, öğretmen adaylarının başarı testi geliştirme yeterliliklerinin incelenmesi olarak belirlenmiştir.

Yöntem

Araştırmanın Modeli

Bu araştırmada öğretmen adaylarının başarı testi geliştirme yeterliliklerinin değerlendirilmesi amaçlandığından, araştırma betimsel türden modellenmiştir (Karasar, 2012). Betimsel araştırmalar, herhangi bir müdahale gerektirmeyen ve var olan durumun olduğu gibi ortaya koymasını amaçlayan kesitsel araştırmalardır (Şata, 2020). Bu araştırmada öğretmen adayların hali hazırdaki soru yazma becerilerinin değerlendirilmesi amaçlandığından bu yönüyle betimsel araştırmaya girmektedir.

Çalışma Grubu

Çalışma grubu, doğu Anadolu bölgesinde bulunan bir üniversitenin eğitim fakültesindeki ilköğretim matematik bölümündeki 57 öğretmen adayından oluşmaktadır. Araştırmanın katılımcılarından 2018-2019 eğitim yılının bahar yarıyılında veriler toplanmıştır. Öğretmen adayları ölçme ve değerlendirme dersini tamamladıktan hemen sonra başarı testi geliştirme sürecine geçmiştir. Araştırmada öğretmen adayları ikiye ve üçer kişilik gruplar oluşturmuş (toplam 24 grup) ve başarı testleri bu gruplar tarafından geliştirilmiştir. Öğretmen adaylarının geliştirmiş oldukları başarı testleri araştırmacı tarafından üç kez puanlamıştır. Puanlama süreci iki aylık sürede gerçekleştirilmiştir. Araştırmada tek bir puanlayıcının farklı zamanlarda değerlendirme yapmasından dolayı sadece puanlayıcı içi güvenirlik katsayısı hesaplanmıştır.

Veri Toplama Aracı

Araştırma kapsamındaki veri, araştırmacı tarafından geliştirilen ve başarı testi geliştirme yeterliliğinin değerlendirilmesini amaçlayan analitik dereceli puanlama anahtarı ile toplanmıştır. Analitik dereceli puanlama anahtarı 6 ölçütten (testin amacı, belirtke tablosunu kullanma, soruların yazılması, uzman görüşü, madde analizi ve madde seçimi, raporlama ve nihai form) ve beşli kategoriye (1 = çok yetersiz, 2 = yetersiz, 3 = orta düzey, 4 = yeterli, 5 = mükemmel) sahip

bir ölçme aracıdır (bakınız Ek 1). Ölçme aracından elde edilen ölçümlerin güvenilirliği Cronbach alfa katsayısı hesaplanmış ve .834 olarak kestirilmiştir. Ölçümlerin geçerliğine kanıt sağlamak amacıyla uzman görüşü (kapsam geçerliği) ve açımlayıcı faktör analizi (yapı geçerliği) yapılmıştır. Kapsam geçerliğine kanıt sağlamak amacıyla ölçme ve değerlendirme alanında doktora derecesine sahip beş uzmandan görüş alınmış ve kapsam geçerlik oranları hesaplanmıştır. Kapsam geçerlik oranları Lawshe tekniği kullanılarak hesaplanmıştır. Alan uzmanları tüm ölçütlerin gerekli olduğuna yönelik görüş belirttiğinden herhangi bir düzeltme yapılmadan ölçme aracına son hali verilmiş ve uygulamaya geçilmiştir. Açımlayıcı faktör analizine ilişkin bulgular ise veri analizi kısmında belirtilmiştir.

Veri Analizi

Veri analizinde çok yüzeyli Rasch ölçme modeli kullanılmıştır. Çok yüzeyli Rasch ölçme modelinin varsayımları olan tek boyutluluk, yerel bağımsızlık ve model veri uyumu test edilmiş ve varsayımların sağlandığı tespit edilmiştir. Analizler tamamen çaprazlanmış desen altında yürütülmüştür. AFA yapılmadan önce varsayımları test edilmiştir (İlgili veri için KMO = .686; Barlett testi χ^2 (sd) = 91.428 (15); $p < .05$). Yapılan açımlayıcı faktör analizi sonucunda ölçme aracının tek boyutlu olduğu bulunmuştur (ilgili veri için faktör yükleri sırayla; .732; .760; .738; .566; .815 ve .761, açıklanan varyans %44.76). Tek boyutluluğun sağlanması yerel bağımsızlığın sağlandığına işaret ettiğinden bu varsayımın da sağlandığı kabul edilmiştir. Model veri uyumunun sağlanması için ± 2 aralığının dışında kalan standartlaştırılmış artık değerlerin sayısı toplam gözlem sayısının %5'inden fazla olmaması ve ± 3 aralığının dışında kalan standartlaştırılmış artık değerlerin de toplam veri sayısının %1'inden fazla olmaması gerektiği belirtilmiştir (Linacre, 2017). Standartlaştırılmamış artık değerler incelendiğinde, ± 2 aralığında 20 (%4.63) ve ± 3 aralığında ise 5 (%1.15) değer olduğu bulunmuş ve model veri uyumunun kabul edilebilir düzeyde olduğu sonucuna ulaşılmıştır (toplam gözlem sayısı $3 \times 6 \times 24 = 432$).

Bulgular

Öğretmen adaylarının başarı testi geliştirme yeterliklerinin incelenmesinin amaçlandığı bu araştırmada ilk olarak başarı testi geliştiren öğretmen adaylarının yeterlilik düzeyleri incelenmiştir. Grup yüzeyine ilişkin ölçüm raporu tablo 1'de verilmiştir.

Tablo 1

Grup Yüzeyine İlişkin Ölçüm Raporu

Test No	Gözlenen Ortalama	Düzeltilmiş Ortalama	Logit değeri	Standart Hata	Uyum İçi	Uyum Dışı
Grup 15	4.72	4.74	1.61	0.48	0.96	1.40
Grup 21	4.61	4.64	1.20	0.42	0.76	0.91
Grup 16	4.56	4.58	1.03	0.40	1.40	1.20
Grup 02	4.39	4.42	0.60	0.36	0.71	0.66
Grup 03	4.39	4.42	0.60	0.36	2.35	2.65
Grup 13	4.39	4.42	0.60	0.36	0.93	0.88
Grup 24	4.39	4.42	0.60	0.36	1.01	1.01
Grup 19	4.33	4.36	0.47	0.35	1.04	1.39
Grup 09	4.28	4.30	0.35	0.34	1.04	0.99
Grup 11	4.28	4.30	0.35	0.34	1.00	0.96
Grup 17	4.28	4.30	0.35	0.34	1.00	1.43

Grup 22	4.22	4.25	0.24	0.34	0.80	0.76
Grup 07	4.06	4.07	-0.09	0.32	1.42	1.87
Grup 01	4.00	4.02	-0.19	0.32	0.78	0.77
Grup 05	4.00	4.02	-0.19	0.32	0.78	0.80
Grup 23	4.00	4.02	-0.19	0.32	0.95	1.06
Grup 12	3.94	3.96	-0.29	0.31	0.94	0.94
Grup 14	3.94	3.96	-0.29	0.31	0.95	1.12
Grup 20	3.94	3.96	-0.29	0.31	0.92	0.95
Grup 06	3.56	3.56	-0.95	0.30	0.43	0.43
Grup 10	3.56	3.56	-0.95	0.30	1.09	1.10
Grup 18	3.33	3.33	-1.32	0.30	0.74	0.73
Grup 08	3.28	3.28	-1.41	0.30	0.78	0.78
Grup 04	3.00	3.00	-1.87	0.30	1.17	1.16
Ortalama	4.06	4.10	0.00	0.07	1.03	1.08
S (Evren)	0.22	0.24	0.47	0.01	0.39	0.45
S(Örneklem)	0.24	0.26	0.52	0.01	0.43	0.49

Model. Evren : RMSE = 0.34 Düzeltmiş S = 0.75 Ayırma Oranı = 2.19

Ayırma İndeksi = 3.26 Güvenirlik = 0.83

Model. Örneklem: RMSE = 0.34 Düzeltmiş S = 0.77 Ayırma Oranı = 2.25

Ayırma İndeksi = 3.33 Güvenirlik = 0.84

Model. Ki-kare (Sabit etkili) : 138.50 sd = 23 p = .00

Model. Ki-kare (Normal) : 19.80 sd = 22 p = .59

Not. S: Standart Sapma, sd: serbestlik derecesi, RMSE: hata kareleri ortalamasının karekökü

Tablo 1'deki logit ve gözlenen ortalama değerleri incelendiğinde, yeterlilik düzeyi en yüksek grubun 15 numaralı, yeterliliği en düşük grubun ise dört numaralı olduğu görülmektedir. Kurulan modele ilişkin ayırma oranının, ayırma indeksinin ve ayırma indeksinin güvenirliliğinin yüksek olduğu bulunmuştur. Bu değerlerin yüksek olması başarı testini geliştiren öğretmen adayları gruplarının farklı yeterlilik düzeylerine sahip olduğunu göstermektedir. Ayrıca sabit etkili ki kare değerinin anlamlı olması bu farklılığın istatistiksel olarak da anlamlı olduğuna işaret etmektedir. Logit değerlerinin herhangi bir eşik değeri olmadığından logit değerleri kullanılarak t-değerleri elde edilmiştir. Elde edilen t-değerleri kritik t değerleri ile karşılaştırılmış ve anlamlı bulunan sonuçlar belirlenmiştir ($t_{kritik}(sd) = 2.064 (23)$; $p = .05$). Başarı testini geliştiren 24 öğretmen adayları grubundan diğer gruplara göre istatistiksel olarak anlamlı olan (düşük ve yüksek yeterlilik) gruplara ilişkin t-değerleri tablo 2'de verilmiştir.

Tablo 2

Başarı Testini Geliştiren Gruplardan Anlamlı Olanlara İlişkin t-değerleri

Grup No	t-değeri	Grup No	t-değeri
4	-6.233	15	3.354
6	-3.167	16	2.575
8	-4.700	18	-4.400
10	-3.167	21	2.857

Tablo 2 incelendiğinde, başarı testini geliştiren öğretmen adayları gruplarına ilişkin t-değerlerinin -6.233 ile 3.354 aralığında değiştiği görülmektedir. Bu değerlere göre değerlendirmesi yapılan 24 grup içinde yeterlilik düzeyi en düşük olan 4.grup iken en yüksek olan ise 15.grup olduğu görülmektedir. Elde edilen t-değerleri incelendiğinde başarı testini geliştiren 24 gruptan beş

tanisinin düşük yeterlilik düzeyine üç tanesinin ise diğerlerine göre daha yüksek yeterlilik düzeyine sahip oldukları bulunmuştur.

Başarı testini geliştiren her bir grubun yeterlilik düzeyi incelendikten sonra analitik dereceli puanlama anahtarındaki ölçütler incelenmiş ve öğretmen adaylarının ölçütlerden hangisinde daha iyi veya kötü oldukları belirlenmeye çalışılmıştır. Buna göre ölçüt yüzeyine ilişkin elde edilen ölçüm raporu tablo 3’de verilmiştir.

Tablo 3

Ölçüt Yüzeyine İlişkin Ölçüm Raporu

Ölçüt Adı	Gözlenen Ortalama	Düzeltilmiş Ortalama	Logit değeri	Standart Hata	Uyum İçi	Uyum Dışı
1. Amaç	4.44	4.51	0.84	0.19	1.70	1.90
2. Belirtke tablosu	3.99	4.02	-0.17	0.16	0.86	0.87
3. Soru yazımı	3.85	3.87	-0.43	0.16	0.73	0.73
4. Uzman görüşü	4.28	4.34	0.43	0.18	1.40	1.43
5. Madde analizi	3.93	3.96	-0.28	0.16	0.96	0.96
6. Raporlama	3.88	3.90	-0.38	0.16	0.56	0.59
Ortalama	4.06	4.10	0.00	0.07	1.03	1.08
S (Evren)	0.22	0.24	0.47	0.01	0.39	0.45
S (Örnekleme)	0.24	0.26	0.52	0.01	0.43	0.49

Model. Evren : RMSE = 0.17 Düzeltilmiş S = 0.44 Ayırma Oranı = 2.59

Ayırma İndeksi = 3.79 Güvenirlilik = 0.87

Model. Örnekleme: RMSE = 0.17 Düzeltilmiş S = 0.49 Ayırma Oranı = 2.88

Ayırma İndeksi = 4.17 Güvenirlilik = 0.89

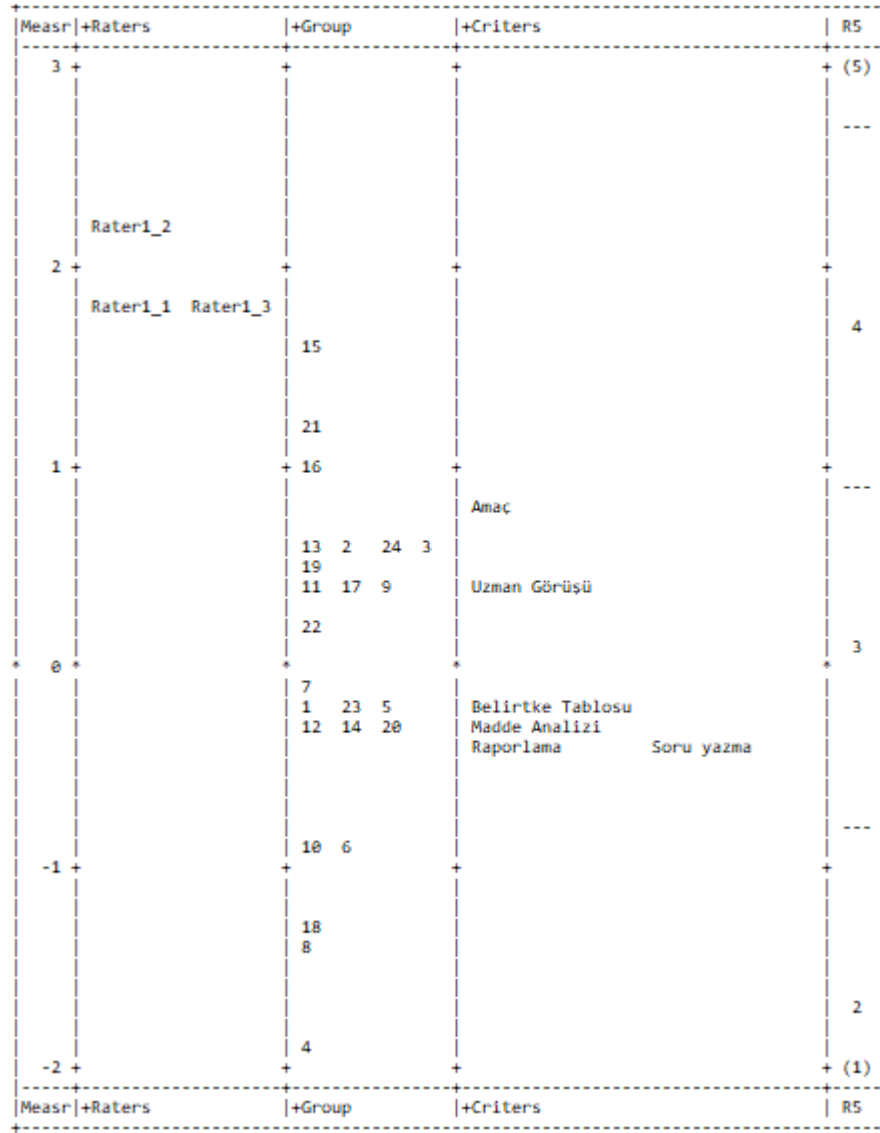
Model. Ki-kare (Sabit etkili) : 41.10 sd = 5 p = .00

Model. Ki-kare (Normal) : 4.50 sd = 4 p = .35

Not. S: Standart Sapma, sd: serbestlik derecesi, RMSE: hata kareleri ortalamasının karekökü

Tablo 3 incelendiğinde, öğretmen adaylarının başarı testi geliştirme sürecinde yeterlilik düzeylerinin en düşük olduğu veya en çok zorlandıkları ölçütün anlaşılır soru yazımı (logit = -.43) olduğu görülmektedir. Diğer yandan öğretmen adaylarının başarı testi geliştirme sürecinde yeterlilik düzeylerinin en yüksek olduğu veya en az zorlandıkları ölçütün ise başarı testinin amacını belirleme (logit = .84) olduğu görülmektedir. Kurulan modele ilişkin ayırma oranı, ayırma indeksi ve ayırma indeksinin güvenirliliğinin yüksek olduğu görülmektedir. Bu değerlerin yüksek olması, ölçütlerin istatistiksel olarak birbirinden farklı yeterlilik düzeylerine sahip olduğunu göstermektedir. Ayrıca ki kare değerine ilişkin anlamlılık testinin anlamlı olması ($p < .05$) ölçütlerin yeterlilik düzeylerinin istatistiksel olarak farklı olduğunu göstermektedir.

Araştırma kapsamındaki tüm yüzeylerin tek bir ölçek üzerinde gösterildiği değişken haritası şekil 1’de verilmiştir.



Şekil 1. Tüm yüzeylere ilişkin değişken haritası.

Şekil 1 incelendiğinde, puanlayıcının her üç puanlamada birbirine yakın puanlamalar yaptığı görülmektedir. Ayrıca yeterlilik düzeyi en yüksek olan başarı testinin 15, en düşük ise 4 olduğu görülmektedir. Ayrıca öğretmen adaylarının başarı testi geliştirme sürecinde en kolay yapabildikleri testin amacını belirlemek iken en zorlandıkları ölçütün ise anlaşılır soru yazımı olduğu görülmektedir.

Tartışma, Sonuç ve Öneriler

Araştırma kapsamında, 57 ilköğretim matematik öğretmenin hazırlamış oldukları 24 başarı testine ilişkin yeterlilik düzeyleri incelenmiş ve 24 grup içinde beş grubun düşük yeterlilikte üç grubun ise yüksek yeterlilikte oldukları belirlenmiştir. Alanyazın incelendiğinde, öğretmenler ve öğretmen adaylarının ölçme ve değerlendirme faaliyetlerine ilişkin yeterlilik düzeylerinin düşük olduğu görülmektedir (Gaitas ve Alves-Martins, 2017; Şenel ve ark., 2018). Ayrıca öğretmen ve öğretmen adayları ölçme araçlarının hazırlanması, uygulanması ve durum belirlemesinin yapılmasının çok zaman alacağını düşündükleri için başarı testi hazırlamaktan kaçınmaktadırlar (Çakan, 2004; Topkaya ve Yılar, 2016). Ulusal çaptaki geniş ölçekli sınavların sadece çoktan seçmeli maddelerden oluşması, öğretmen ve öğretmen adaylarının sadece bu

soru tipine yönelmelerine neden olmaktadır (Bayat ve Şentürk, 2015). Bu bağlamda mevcut araştırma kapsamında elde edilen sonuçların alanyazında desteklendiği görülmektedir. Öğretmen adaylarının aldıkları lisans eğitiminin aynı olması ve tüm grupların çoktan seçmeli madde formatına göre başarı testi geliştirmiş olmasına rağmen, bazı grupların daha yetersiz olmasının temel nedenlerinden biri her bir grubun farklı kazanımlara yönelik başarı testini geliştirmesinden kaynaklandığı düşünülmektedir.

Ayrıca yeterlilik düzeyleri düşük olan grupların hazırlamış oldukları başarı testlerine ilişkin kazanımlarının üst düzey bilişsel düşünme becerisine sahip kazanımlara sahip olduğu tespit edilmiştir. Alanyazında üst düzey bilişsel becerilere yönelik çoktan seçmeli maddelerin yazımının zor olduğu belirtilmektedir (Kutlu ve ark., 2014; Moore, 2009). Bu bağlamda öğretmen adaylarının da bu nedenden dolayı başarı testi hazırlama yeterlilikleri olması gerekenden daha düşük olduğu düşünülmektedir. Bu açıdan bakıldığında farklı kazanımlara yönelik başarı testinin geliştirilmesi ve değerlendirilmesinin yapılması çalışmanın sınırlılığı olarak kabul edilmektedir. Ayrıca öğretmen adaylarının daha önce çoktan seçmeli madde yazımı ve başarı testi geliştirmemesinden dolayı üst düzey zihinsel becerilere yönelik kazanımları soru formatına dönüştürmede zorluk çektikleri düşünülmektedir.

Araştırma kapsamında ulaşılan diğer bir sonuç ise öğretmen adaylarının başarı testi geliştirme sürecinin hangi aşamasında zorlandıkları belirlenmiştir. Başarı testinin geliştirilme süreci dikkate alınarak hazırlanan analitik dereceli puanlama anahtarının ölçütleri incelendiğinde, öğretmen adaylarının en fazla zorlandıkları süreç kazanıma uygun ve anlaşılır madde yazımı olduğu belirlenmiştir. Bu sonuç bir önceki sonucun uzantısı olarak ortaya çıkmış ve öğretmen adaylarının üst düzey bilişsel beceriye yönelik soru yazamamalarından kaynaklandığı belirlenmiştir. Çetinkaya (2019) tarafından yapılan araştırmada da sosyal bilgiler öğretmen adaylarının daha çok alt düzey bilişsel becerilere yönelik soru hazırladıkları ve üst düzey bilişsel becerileri ölçen maddeleri yazamadıklarını belirlemiştir. Değişken haritasında ölçütlere ilişkin yüzey incelendiğinde, başarı testinin ilk aşaması olan amaç belirleme ölçütünün diğer ölçütlere göre daha yüksek bir yeterliliğe sahip olduğu görülmektedir. Daha sonra ise uzman görüşü almanın geldiği görülmekte fakat diğer dört ölçütü yapabilme yeterliliklerinin çok düşük olduğu tespit edilmiştir. Bu durum öğretmen adaylarının başarı testinin birçok aşamasında düşük yeterliliğe sahip olduklarına işaret etmektedir. Öğretmen adaylarının soru yazma ve başarı testi geliştirme yeterliliklerinin düşük olması, bu ölçme araçları ile elde edilen ölçümlerin geçerliği ve güvenilirliğinin düşük olmasına neden olmaktadır. Alanyazında yapılan farklı çalışmalarda da öğretmenlerin başarı testi geliştirme süreçlerinde düşük yeterliliğe sahip oldukları bildirilmiştir (Bay, 2015; Güneyli ve Abbasoğlu, 2015).

Sonuç olarak, başarı testi geliştiren ilköğretim matematik aday öğretmenlerinin farklı yeterlilik düzeylerine sahip oldukları ve genel olarak üst düzey kazanımları ölçen madde yazmadıkları tespit edilmiştir. Başarı testinin geliştirilmesi sürecinde en fazla zorlandıkları ölçütün kazanıma uygun madde yazımı olduğu ve en az zorlandıkları ölçütün ise başarı testinin amacının belirlenmesi olduğu bulunmuştur.

Bu sonuçlara dayalı olarak eğitim fakültelerinde okutulan ölçme ve değerlendirme dersinin mevcut programının aksine haftalık ders saatinin en az iki katına çıkartılması ve uygulama boyutunun da teorik boyut kadar olması gerektiği düşünülmektedir. Ayrıca soru yazma

eğitimlerinin öğretmen adaylarına verilmesi yapılacak ölçümlerin daha geçerli ve güvenilir olmasına katkı sağlayacağı beklenmektedir.

Çıkar Çatışması Bildirimi

Yazar, bu makalenin araştırılması, yazarlığı ve/veya yayınlanmasına ilişkin herhangi bir potansiyel çıkar çatışması beyan etmemiştir.

Destek/Finansman Bilgileri

Yazar, bu makalenin araştırılması, yazarlığı ve / veya yayınlanması için herhangi bir finansal destek almamıştır.

Etik Kurul Kararı

Araştırma, 2018-2019 eğitim yılının bahar yarıyılında gerçekleştiğinden, herhangi bir etik kurul izni alınmamıştır.

Kaynakça

- Açıkgöz, M. & Karslı, F. (2015). Alternatif ölçme-değerlendirme teknikleri kullanılarak iş ve enerji konusunda geliştirilen başarı testinin geçerlilik ve güvenilirlik analizi. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 4(1), 1-25.
- Akbulut, H. İ. & Çepni, S. (2013). Bir üniteye yönelik başarı testi nasıl geliştirilir? İlköğretim 7. sınıf kuvvet ve hareket ünitesi. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 2(1), 18-44.
- Altun, T. & Gülay, A. (2017). Göreve yeni başlayan öğretmenlerin yeterlik algılarının ve karşılaştıkları sorunların belirlenmesi. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, (31), 738-749. <https://doi.org/10.14582/DUZGEF.1837>
- Bay, D. N. (2015). The question asking skills of preschool teacher candidates: Turkey and America example. *Journal of Education and Training Studies*, 4(1), 161-169. <https://doi.org/10.11114/jets.v4i1.1141>
- Bayat, S. ve Şentürk, Ş. (2015). Fizik, Kimya, Biyoloji ortaöğretim alan öğretmenlerinin alternatif ölçme değerlendirme tekniklerine ilişkin görüşleri. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 4(1), 118-135.
- Boud, D. (1995, Eylül, 24-27). *Ensuring that assessment contributes to learning* [Oral presentation]. International confrence on problem-based learning in higher education, Linköping, Sweden.
- Boyras, A. (2007). *İlköğretim okullarında görev yapan aday öğretmenlerin sınıfta karşılaştıkları disiplin sorunları* [Yüksek lisans tezi, Gazi Üniversitesi]. <https://tez.yok.gov.tr/>
- Çakan, M. (2004). Öğretmenlerin ölçme-değerlendirme uygulamaları ve yeterlik düzeyleri: İlk ve ortaöğretim. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 37(2), 99-114. https://doi.org/10.1501/Egifak_0000000101
- Çetinkaya, B. (2019). Sosyal bilgiler öğretmenliği öğrencilerinin bloom taksonomisine göre soru yazmaya ilişkin başarı düzeylerinin değerlendirilmesi. *International Journal of Social Science*, 2(1), 115-148.
- Gaitas, S. & Alves-Martins, M. (2017). Teacher perceived difficulty in implementing differentiated instructional strategies in primary school. *International Journal of Inclusive Education*, 21(5), 544-556. <https://doi.org/10.1080/13603116.2016.1223180>
- Güler, N. & Gelbal, S. (2010). Açık uçlu matematik sorularının güvenilirliğinin klasik test kuramı ve genellenabilirlik kuramına göre incelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri*, 10(2), 991-1019.
- Güneyli, A. & Abbasoğlu, Ş. (2015). Türkçe dersi yazılı sınav sorularına ilişkin değerlendirme: Kıbrıs örneği. *Journal of Academic Studies*, 17(67), 53-76.
- İnceçam, B., Demir, E. & Demir, E. (2018). Ortaokul öğretmenlerinin sınıf içi ölçme ve değerlendirmelerde yazılı yoklamalarda kullandıkları açık uçlu maddeleri hazırlama yeterlikleri. *İlköğretim Online*, 17(4), 1912-1927. <https://doi.org/10.17051/ilkonline.2019.506900>
- Karaduman, B. & Yelken, T. Y. (2020). Öğretmen adaylarının değerlendirme tercihleri ile ölçme ve değerlendirme genel yeterlik algılarının belirlenmesi. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 29(1), 339-353. <https://doi.org/10.35379/cusosbil.651030>
- Karasar, N. (2012). *Bilimsel araştırma yöntemi* (17.Baskı). Nobel Yayıncılık.
- Korkmaz, İ., Saban, A. & Akbaşlı, S. (2004). Göreve yeni başlayan sınıf öğretmenlerinin karşılaştıkları güçlükler. *Kuram ve Uygulamada Eğitim Yönetimi*, 38, 266-277.
- Kutlu, Ö., Doğan, C.D. & Karakaya, İ. (2014). *Öğrenci başarısının belirlenmesi: Performansa ve portfolyoya dayalı durum belirleme* (4.Baskı). Pegem Akademi Yayıncılık.
- Linacre, J. M. (2017). *FACETS* (Version 3.80.0) [Computer Software]. Chicago, IL: MESA Press.

- Moore, B.B. (2009). *Consideration of rater effects and rater design via signal detection theory* [Doktora Tezi, Columbia University]. <http://www.proquest.com/>
- Smith, L. & Southerland, S. A. (2007). Reforming practice or modifying reforms? Elementary teachers' response to the tools of reform. *Journal of Research in Science Teaching*, 44(3), 396-423. <https://doi.org/10.1002/tea.20165>
- Şata, M. (2020). Nicel araştırma yaklaşımları. E. Oğuz. (Ed.), *Eğitimde araştırma yöntemleri* (1.baskı) içinde (s. 77-90). Eğiten Kitap Yayınları.
- Şenel, S., Pekdağ, B. & Günaydın, S. (2018). Kimya öğretmenlerinin eğitimde ölçme ve değerlendirme süreçlerinde yaşadıkları problemler ve yetersizlikler. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, 12(1), 419-441. <https://doi.org/10.17522/balikesirnef.437824>
- Topkaya, Y. & Yılar, B. (2016). Sosyal Bilgiler Öğretmenlerinin Alternatif Ölçme ve Değerlendirme Teknikleri Hakkındaki Görüşleri. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi*, 18(1), 593-610. <https://doi.org/10.17556/jef.38105>
- Turgut, M. F. & Baykul, Y. (2010). *Eğitimde ölçme değerlendirme* (7. Baskı). Pegem Yayıncılık.

İletişim/Correspondence

Doç. Dr. Mehmet ŞATA
msata@agri.edu.tr