***Investigation of Achievement Tests Used in Doctoral Dissertations Department of Primary Education (2012-2017)***

***Celal BOYRAZ***

*Bayburt University, Faculty of Education*

|  |  |
| --- | --- |
| **Abstract**  *The purpose of this research, it is to examine the appropriateness of achievement tests used in doctoral dissertations department of Primary Education. The research was designed according to the analytical research model. In this context, the doctoral dissertations carried out in the department of Primary Education and using the achievement test as a measurement instrument and accessible was reached from YÖKTEZ database. The study was conducted with the help of a checklist including literature review, item pool, table of specifications, expert opinion, content validity studies, pilot application, item analysis (item discrimination and difficulty) and reliability values. In the scope of the research, achievement tests used as a measurement instrument in 62 doctoral dissertations were examined. As a result of the examination, it was concluded that the achievement test development was inadequate at every stage. These inadequacies are mostly seen preparation table of specifications in the preparatory stage (66.1%), get expert opinion (53.2%) and content validity (88.7%) in the application stage. In the analysis stage, the last step of the achievement test development process, it is noteworthy that the deficiencies are small compared to the other stages. Only 3 (4.8%) of the number of achievement tests in which each level is sufficient indicates that the achievement tests used in the theses do not have the desired quality. Some suggestions were made in the direction of the findings obtained from the research.*  **Keywords:** *doctoral dissertations, achievement tests, department of primary education* | İnönü University  Journal of the Faculty of Education  Vol 19, No 3, 2018  pp. 14-28  DOI: 10.17679/inuefd.327321 |
| Received : 08.08.2017  Accepted : 24.09.2018 |

**Suggested Citation**

Boyraz, C. (2018). Investigation of achievement tests used in doctoral dissertations department of primary education (2012-2017). *Inonu University Journal of the Faculty of Education*, 19(3), 14-28. DOI: 10.17679/inuefd.327321

**EXTENDED ABSTRACT**

**Introduction**

The researches conducted in the field of education is very important in terms of shaping their educational practices and educational policies. It is expected that the researches that can lead education policies are qualified in every sense. The patterns used in the research, measurement tools, correct conducting of the research process, making accurate data analysis are factors affecting the quality of the research. Especially by the development process of measurement tools used reliable and validity or not it will directly affect the outcome of the study. It is an important obstacle to reaching the right results when researching the data obtained from measurement instruments which are not fully complied with the development process of measurement tool and whose validity and reliability are not sufficient. One of these measurement tools, which require a considerably long process and labour to develop, is achievement tests. Achievement tests are often used in educational research to determine the functioning of the program and to determine the goals for which learners have difficulty learning, and to determine the level of learning of students before they begin their teaching program. These studies are mostly experimental research designs adopted. The effectiveness of the treatment process is checked by comparing the scores obtained by the students before and after the training process through the achievement tests in the educational researches that have adopted the experimental design. In this context, qualified the achievement tests directly affect the results of the research. In order to provide this quality, the development process of achievement test has been tried to be explained in seven stages. These stages; determining the object preparing achievement test, determination of the scope of the test, formation of table of specification, preparation of item pool and taking the expert opinion, preparation and application of the pilot test, doing item analysis and finalize the achievement test. It is expected that detailed explanations of these steps be carefully followed in the research reports using the success test. Because both researchers who study the results of the research and stakeholders who guide the educational policies will decide on the nature of the means of measuring that the results of the research are valid and reliable. When held in Turkey in graduate thesis researches revealed that the thought of heavy methodological error, as is conceivable that there may be serious bugs in the development of measurement tools, such as achievement test. However, it did not find any research when developing educational achievement tests used as a measuring tool to research what extent that these steps be taken.

***Purpose***

The purpose of this research, is to examine the appropriateness of achievement tests used in doctoral dissertations department of Primary Education.

***Method***

The research was designed according to the analytical research model, which is outside the qualitative and quantitative research paradigms. The analytical research model is a research method in which documents, documents, records and other media are analysed in terms of events, thoughts, concepts and works. In this context, the doctoral theses carried out in the department of Primary Education and using the achievement test as a measurement instrument and accessible was reached from YÖKTEZ database. The study was conducted with the help of a checklist including literature review, item pool, table of specifications, expert opinion, content validity studies, pilot application, item analysis (item discrimination and difficulty) and reliability values. Descriptive analysis was carried out by quantification based on the categories determined for the data obtained in the research. In the scope of the research, achievement tests used as a measurement instrument in 62 doctoral dissertations were examined. Descriptive analysis was carried out by quantification based on the categories determined for the data obtained in the research.

***Findings***

Because of the examination, it was concluded that the achievement test development was inadequate at every stage. These inadequacies are mostly seen preparation table of specifications in the preparatory stage (66.1%), get expert opinion (53.2%) and content validity (88.7%) in the application stage. In the analysis stage, the last step of the achievement test development process, it is noteworthy that the deficiencies are small compared to the other stages. Only 3 (4.8%) of the number of achievement tests in which each level is sufficient indicates that the achievement tests used in the theses do not have the desired quality.

***Discussion & Conclusion***

In the scope of the research, success tests used as a measurement tool in 62 doctoral dissertations were examined. As a result of the examination, it is concluded that the success test development is inadequate at every stage. These inadequacies are mostly seen preparation table of specifications in the preparatory stage (66.1%), get expert opinion (53.2%) and content validity (88.7%) in the application stage. In the analysis stage, the last step of the achievement test development process, it is noteworthy that the deficiencies are small compared to the other stages. Achievement test development requires a holistic process. Only 3 (4.8%) of the number of achievement tests in which each level is sufficient indicates that the achievement tests used in the theses do not have the desired quality. This result is similar the result of other research conducted within the scope of qualitative inquiries of educational researches (Boyraz ve Serin, 2016; Evrekli, İnel, Deniş ve Balım, 2011; Kabaca ve Erdoğan, 2007; Karadağ, 2010; Konan ve Oğuz, 2015; Mutluer ve Yandı, 2012; Şahin, Calp, Bulut ve Kuşdemir, 2013). Of course, evaluating the achievement tests examined by a single investigator is one of the most important limitations of the research. Some suggestions were made in the direction of the findings obtained from the research.

***İlköğretim Anabilim Dalı Doktora Tezlerinde Kullanılan Başarı Testlerinin İncelenmesi (2012-2017)***

***Celal BOYRAZ***

*Bayburt Üniversitesi, Eğitim Fakültesi*

|  |  |
| --- | --- |
| **Öz**  *Bu araştırmanın amacı, İlköğretim anabilim dalında hazırlanan doktora tezlerinde kullanılan başarı testlerinin, başarı testi geliştirme adımlarına uygunluğunu incelemektir. Araştırma, analitik araştırma modeline göre desenlenmiştir. Bu bağlamda İlköğretim anabilim dalında 2012-2017 yılları arasında hazırlanan veri toplama aracı olarak başarı testi kullanılan ve erişimine izin verilen 67 doktora tezine YÖKTEZ veri tabanından ulaşılmıştır. Bu tezlerde yer alan başarı testleri; alanyazın taraması, madde havuzu, belirtke tablosu, uzman görüşü, kapsam geçerliği çalışmaları, pilot uygulama, madde analizleri (ayırt edicilik, güçlük) ve güvenirlik değerlerinin yer aldığı araştırmacı tarafından geliştirilen bir dereceli puanlama anahtarı yardımıyla incelenmiştir. Araştırma kapsamında toplam 62 doktora tezinde ölçme aracı olarak kullanılan başarı testleri incelenmiştir. İnceleme sonucunda başarı testi geliştirmenin her aşamasında yetersizlikler olduğu sonucuna ulaşılmıştır. Bu yetersizlikler en çok hazırlık aşamasında belirtke tablolarının hazırlanmasında (%66,1), uygulama aşamasında uzman görüşü alma (%53,2) ve kapsam geçerliği çalışmalarında (%88,7) görülmektedir. Başarı testi geliştirme sürecinin son basamağı olan analiz aşamasında ise eksikliklerin diğer aşamalara göre az olması dikkati çeken bir bulgudur. Her bir aşamanın yeterli düzeyde olduğu başarı testi sayısının sadece 3 (%4,8) olması tezlerde kullanılan başarı testlerinin istenilen niteliğe sahip olmadığını göstermektedir. Araştırmadan elde edilen bulgular doğrultusunda bazı önerilerde bulunulmuştur.*  **Anahtar Kelimeler:** *doktora tezi, başarı testleri, ilköğretim anabilim dalı* | İnönü Üniversitesi  Eğitim Fakültesi Dergisi  Cilt 19, Sayı 3, 2018  ss. 14-28  DOI: 10.17679/inuefd.327321 |
| Gönderim Tarihi : 08.08.2017  Kabul Tarihi : 24.09.2018 |

**Önerilen Atıf**

Boyraz, C. (2018). İlköğretim anabilim dalı doktora tezlerinde kullanılan başarı testlerinin incelenmesi. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, 19(3), 14-28. DOI: 10.17679/inuefd.327321

**GİRİŞ**

Eğitim alanında gerçekleştirilen araştırmaların sayısında her geçen yıl büyük bir artış olduğu gözlemlenmektedir. Bu gözlemi eğitim bilimleri kapsamında son 10 yılda hazırlanan lisansüstü tez sayıları da desteklemektedir. Örneğin; YÖKTEZ veri tabanında Eğitim Bilimleri Enstitüsü bünyesinde 2007-2017 yılları arasında 14.159 (31.05.2017 itibariyle) kayıt bulunmaktadır. Bu kayıtların 5.051 adedi 2007-2011 yılları arasında, 9.108 adedi de 2012-2017 yılları arasında hazırlanmıştır. Eğitim araştırmalarının eğitim uygulamalarına ve eğitim politikalarına yön verme açısından önemli bir potansiyeli bulunduğundan araştırmalardaki niceliksel artış olumlu bir gelişme olarak düşünülebilir. Bununla birlikte eğitim uygulamaları ve politikalarına yön verebilecek araştırmaların her anlamda nitelikli olması beklenir.

Eğitim araştırmalarında niteliğin sorgulanması, araştırmalarda ulaşılan sonuçların kalitesini ve kullanılabilirliğini ortaya koyması açısından büyük bir önem taşımaktadır (Karadağ, 2009). Ancak eğitim bilimleri alanında yapılmış araştırmaları nitelik açısından inceleyen araştırmalar, bu alanda ciddi hataların yapıldığını göstermektedir. Yapılan nitelik sorgulama araştırmalarına bakıldığında çoğunlukla yöntemsel ve istatistiksel bağlamlarda ele alınan araştırmalar olduğu söylenebilir (Boyraz ve Serin, 2016; Evrekli, İnel, Deniş ve Balım, 2011; Kabaca ve Erdoğan, 2007; Karadağ, 2010; Konan ve Oğuz, 2015; Şahin, Calp, Bulut ve Kuşdemir, 2013). Bu araştırmalarda, benimsenen desene uygunluk, veri toplama araçlarının geliştirilme süreçleri, araştırma sürecinin doğru yönetilmesi, doğru veri analizlerinin yapılması gibi başlıklarda incelemeler yapılmıştır. Bu başlıklar incelendiğinde, kullanılan veri toplama araçlarının geçerli ve güvenilir olması/olmaması araştırma sonucunu doğrudan etkileyeceği için oldukça önemli görülmektedir.

Veri toplama aracı geliştirme süreçlerine harfiyen uyulmayan, geçerliği ve güvenilirliği yeterli olmayan ölçme araçlarından elde edilecek veriler araştırmada doğru ve tutarlı sonuçlara ulaşılmasında önemli bir engeldir. Araştırmalarda kullanılan veri toplama araçlarının ölçme aracı geliştirme/uyarlama süreçlerine uygunluğunu konu edinen çalışmalar alanyazında önemli bir yer tutmaktadır. Bu çalışmaların büyük çoğunluğu likert tipi ölçek geliştiren çalışmaların incelenmesini esas almaktadır (Bayık ve Gürbüz, 2016; Çüm ve Koç, 2013; Delice ve Ergene, 2015; Erkuş, 2000; Erkuş, 2007; Erkuş, 2016; Şahin ve Boztunç Öztürk, 2018; Karadağ, 2011; Mor Dirlik, 2014; Öztürk, Eroğlu ve Kelecioğlu, 2015). Karadağ (2011) yaptığı çalışmasında likert tipi ölçeklere ek olarak araştırmalarda kullanılan görüşme formu, gözlem formu ve başarı testi gibi ölçme araçlarının da nitelik düzeylerini incelemiştir. Likert tipi ölçeklerin ölçme aracı geliştirme adımlarına uygunluğunu inceleyen çok sayıda araştırma varken, tıpkı ölçekler gibi geliştirilmesi oldukça uzun bir süreç ve emek gerektiren başarı testlerini bu açıdan ele alan yeterli sayıda çalışma bulunmamaktadır. Bu eksiklik elde ki araştırmanın gerçekleştirilmesine önemli bir neden oluşturmaktadır.

Alanyazında güç testleri, maksimum performans testleri, bilgi testleri olarak da isimlendirilen başarı testleri çoğunlukla deneysel desen benimsenen araştırmalarda kullanılmaktadır. Deneysel desen benimsenen eğitim araştırmalarında, başarı testleri yoluyla öğrencilerin denel işlem öncesinde ve sonrasında elde ettiği puanlar karşılaştırılarak denel işlemin etkililiği kontrol edilmektedir. Başarı testleri, öğrencilerin öğretim programına başlamadan önce giriş davranışlarını tespit etmede, programın işlerliğini ve öğrencilerin öğrenme güçlüğü yaşadığı hedefleri belirlemede ve öğrencilerin öğrenme düzeylerini saptamada sıklıkla eğitim araştırmalarında kullanılmaktadır (Adıgüzel, 2016; Fraenkel, Wallen ve Hyun, 2012; Tan, 2006). Bu testler hedef kitlenin bilişsel düzeyleri hakkında nicel veri sağlamasıyla birlikte, bireylerin konu alanlarına yönelik ihtiyaçlarını gösterdiği için eğitim programlarının geliştirilmesinde de önemli bir işleve sahiptir (Adıgüzel, 2016). Bu bağlamda veri toplama aracı olarak başarı testi kullanılan araştırmaların doğru sonuçlara ulaşabilmesi için başarı testlerinin geçerli ve güvenilir araçlar olarak tasarlanması gerekir. Dolayısıyla araştırmalarda kullanılacak başarı testlerinin, başarı testi hazırlama ilkelerine uygun olarak hazırlanması oldukça önemlidir.

Geçerli ve güvenilir başarı testleri hazırlayabilmek için başarı testi geliştirme aşamalarına hassasiyetle uyulması gerekir. Bu aşamalar alanyazında farklı biçimlerde açıklanmaya çalışılmıştır (Adıgüzel, 2016; Anastasi, 1988; Büyüköztürk, Çakmak, Akgün, Karadeniz ve Demirel, 2014; Fraenkel, Wallen ve Hyun, 2012; Turgut ve Baykul, 2010). Örneğin, Anastasi (1988) başarı testi geliştirme sürecini sekiz aşamada ele alırken, Adıgüzel (2016) bu süreci hazırlık, uygulama ve analiz olmak üzere üç aşamada ele almıştır. Ancak genel olarak bakıldığında başarı testi geliştirmek için şu ortak aşamalardan söz edilebilir.

• Başarı testinin amacının belirlenmesi,

• Testin kapsamının belirlenmesi,

• Belirtke tablosunun oluşturulması,

• Madde havuzunun hazırlanıp uzman görüşüne başvurularak kapsam geçerliği çalışmalarının yapılması (Kapsam geçerliği ortalaması (KGO) ve Kapsam geçerliği indeksi (KGİ) hesaplamaları),

• Deneme formu hazırlanarak pilot uygulama yapılması,

• Madde analizleri yapılarak teste son halinin verilmesi.

Başarı testi kullanılan araştırma raporlarında bu aşamaların özenle takip edildiğinin detaylı bir biçimde açıklanması beklenmektedir. Çünkü çalışma sonuçlarını inceleyen gerek araştırmacılar ve uygulayıcılar gerekse eğitim politikalarına yön veren paydaşlar, çalışma sonuçlarının geçerli ve güvenilir olduğuna ölçme aracının niteliği doğrultusunda karar vereceklerdir. Türkiye’de yapılan lisansüstü tezlerde ağır metodolojik hataların olduğunu ortaya koyan araştırmalar düşünüldüğünde, başarı testlerinin geliştirilmesinde test hazırlama sürecinde dikkate alınması gereken ilkelere ne derece uyulduğu cevaplanmayı bekleyen bir soru olarak karşımıza çıkmaktadır. Öyle ki ölçme aracı olarak kullanılan psikolojik testlerin ve ölçeklerin geliştirme/uyarlama sürecinde yapılan hatalar birçok araştırmaya konu olmuşken, başarı testlerinin geliştirilmesinde test hazırlama sürecinde takip edilmesi gereken adımlara ne ölçüde dikkat edildiğine dair yeterince araştırmaya rastlanmamıştır.

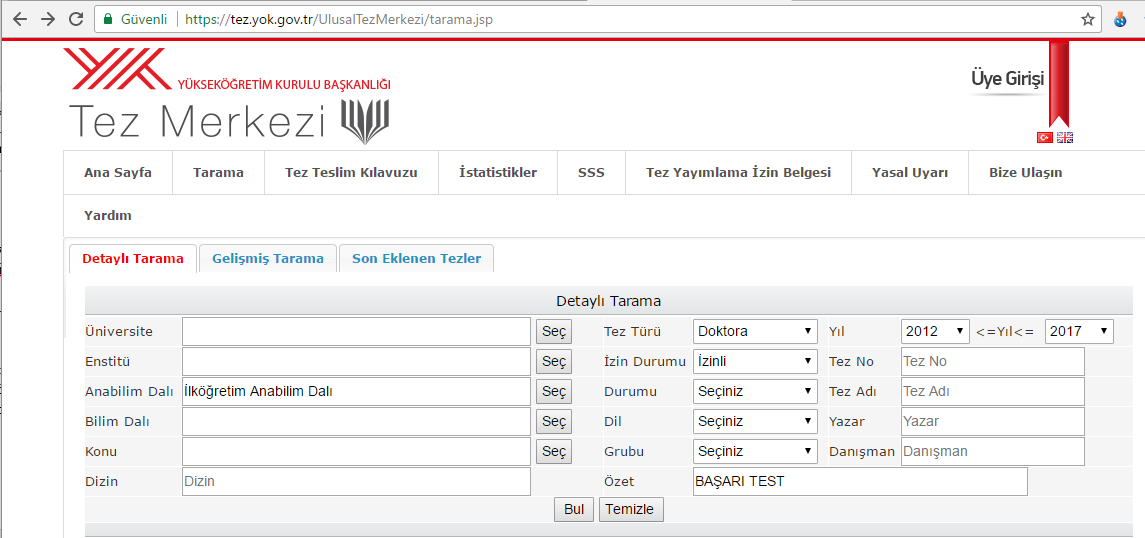
Mutluer ve Yandı (2012) tarafından gerçekleştirilen araştırmada 2010-2012 yılları arasında eğitim bilimleri enstitüleri bünyesinde gerçekleştirilen 50 lisansüstü tez (4 doktora, 46 yüksek lisans) incelenmiş ve sadece 16 tezin başarı testi geliştirme aşamalarına uygun olduğu tespit edilmiştir. Bu araştırma özet bildiri olarak yayınlandığı için araştırma hakkında detaylı bilgilere ulaşılamamıştır. Karadağ (2011) tarafından gerçekleştirilen araştırmada ise 2003-2007 yılları arasında eğitim bilimleri doktora tezlerinde kullanılan 77 başarı testi nitelik açısından incelenmiştir. Çalışmada kullanılan veri toplama aracında 5 maddelik bir rubrik kullanmıştır. Maddeler başarı testinin geliştirme nedenleri, geliştirme sürecinin açıklanması, pilot çalışma süreci, puanlama ile ilgili bilgi verilmesi ve testin araştırma amacına uygunluğu biçimindedir. İnceleme sonucunda başarı testleri niteliksiz seviyesinde değerlendirilmiş ve başarı testlerinin geçerlik ve güvenirliğinin saptanmaması ve testlerin puanlanmasıyla ilgili bilgilere yer verilmemesi başarı testi kullanılan tezlerde en sık yapılan hatalar olarak belirlenmiştir. Yabancı alanyazında ise Hall (1986) ve Onwuegbuzie (2002) tarafından yapılan araştırmalarda da benzer bulgulara ulaşılmıştır. Bu bağlamda eğitim bilimleri alanında yapılan lisansüstü tezlerde kullanılan başarı testlerinin mevcut durumunun resmedilmesinin niteliği artırma adına önemli katkılar getirebileceği düşünülmektedir. Bu düşünceyle araştırmanın amacı, İlköğretim anabilim dalı bünyesinde gerçekleştirilen doktora tezlerinde kullanılan başarı testlerinin, başarı testi geliştirme adımlarına uygunluğunu incelemektir.

## YÖNTEM

Bu araştırmada analitik araştırma yöntemi benimsenmiştir. Analitik araştırma yönteminde dokümanlar, belgeler ve kayıtlar, çeşitli fikirler açısından analiz edilir (Mcmillan, 2004). Çoğu eğitim araştırması nicel ve nitel araştırma olarak sınıflandırılmaktadır. Bu iki sınıflamaya uymayan çalışmalar analitik araştırmalar olarak adlandırılmaktadır. Analitik araştırmalar nitel ve nicel araştırma özelliklerini içerir ve kendi içinde tarihsel analiz, hukuki analiz, kavram analizi ve karma yöntem araştırmaları olarak sınıflandırılır. Araştırmacı, kavramı ya da doğrudan gözlenebilen/gözlenemeyen geçmiş bir olayı anlamak için verileri tanımlar, araştırır ve sonra sentezler. Elde edilen verilerin yorumlanmasında bağlam önemlidir. Doğrudan etkileşim gerektirmeyen analitik araştırma modeli genellikle dokümanlara dayalı veri toplama ve analiz esasına dayalıdır (Burkett, 1990; Mcmillan ve Schumacher, 2014). Bu bağlamda ilköğretim anabilim dalında gerçekleştirilen doktora tezleri doküman olarak kabul edilmiş ve veri toplama aracı olarak doküman incelemesinden faydalanılmıştır.

**Veri Kaynakları**

Araştırma verilerini belirlemek için güncellik ve yöntembilim alanında yaşanan gelişmeler göz önünde bulundurularak 2012-2017 yılları arasında birbirinden farklı disiplinlere hitap eden birçok bilim dalını bünyesinde barındıran ilköğretim anabilim dalında gerçekleştirilmiş doktora tezlerine ulaşılmaya çalışılmıştır. Bu kapsamda öncelikle İlköğretim anabilim dalında gerçekleştirilen ve veri toplama aracı olarak başarı testi kullanılan ve erişimine izin verilen doktora tezlerine YÖKTEZ veri tabanından ulaşılmıştır. Bu tezlerin toplam sayısı 67 olarak tespit edilmiştir (07.06.2017 tarihinde yapılan arama sonucudur). Arama ölçütleri Fotoğraf 1’de gösterilmiştir.



Fotoğraf 1. Başarı testi kullanılan doktora tezlerine ulaşılma ölçütleri

Fotoğraf 1’de görüldüğü gibi üniversitelerde enstitüler bünyesinde yürütülen lisanüstü tez çalışmaları anabilim dalı ve bilim dalı yapılanması biçimindedir. İlköğretim Anabilim Dalı 2017 yılından sonra Temel Eğitim Anabilim Dalı olarak adlandırılmış ve bünyesinde Sınıf Öğretmenliği ve Okul Öncesi Öğretmenliği bilim dalları yer almıştır. Bu nedenle bu çalışmanın kapsamı sadece İlköğretim Anabilim dalı olarak kayda geçen doktora tez çalışmalarıyla sınırlıdır.

**Doktora Tezlerinin Hazırlandığı Alanlar ve Yıllar**

İlköğretim ana bilim dalı bünyesinde 2012-2017 yılları arasında ölçme aracı olarak çoktan seçmeli başarı testleri kullanılarak hazırlanan doktora tezlerinin alanlara ve yıllara göre dağılımı Tablo 1 de sunulmuştur.

Tablo 1.

Doktora tezlerinin alanlara ve yıllara göre dağılımı

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Yıl  Alan | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | Toplam |
| Fen Bilgisi | 5 | 12 | 4 | 4 | 4 | - | 29 |
| Matematik | 1 | 2 | 4 | 1 | 2 | - | 10 |
| Sosyal Bilgiler | 3 | 2 | 2 | 2 | 1 | - | 10 |
| Sınıf Öğretmenliği | 1 | - | 4 | 1 | 1 | 1 | 8 |
| İlköğretim | 1 | 1 | 1 | - | 2 | - | 5 |
| Toplam | 11 | 17 | 15 | 8 | 10 | 1 | 62 |

Tablo 1 incelendiğinde 2012-2017 yılları arasında ilköğretim anabilim dalı bünyesinde ölçme aracı olarak çoktan seçmeli başarı testi kullanılarak gerçekleştirilen doktora tezinin Fen Bilgisi Bilim Dalında (n:29) daha yüksek olduğu görülmektedir. Onu sırasıyla Matematik, Sosyal Bilgiler ve Sınıf Öğretmenliği bilim dalları izlemektedir. Herhangi bir bilim dalı ifadesinin yer almadığı ve ilköğretim bilim dalı olarak adlandırılan tez sayısı ise 5’tir. İlköğretim ana bilim dalı bünyesinde erken çocukluk dönemi çalışmalarının yapıldığı okul öncesi bilim dalı olmasına rağmen alanın özelliğinden dolayı başarı testi kullanılan bir doktora tezine rastlanılmamıştır. Yıllara göre bakıldığında ise 2013 (n:17) ve 2014 (n:15) yıllarında başarı testi kullanılan doktora tezleri sayısı daha yüksektir.

**Başarı Testi Geliştirme Aşamaları Rubriği**

Verilerin elde edilmesinde araştırmacı tarafından hazırlanan “Başarı Testi Geliştirme Aşamaları Rubriği” kullanılmıştır. Bu rubrik hazırlanırken alanyazın taraması yapılmış ve başarı testi geliştirme aşamaları şu şekilde belirlenmiştir.

***Amaç belirleme ve alanyazın taraması****:* Bir başarı testi hazırlanırken ilk olarak testin amacının belirlenerek test ile ilgili alanyazın taraması yapılmalıdır. Bunun için test yoluyla ölçülmek istenenlerin açık ve net olarak ortaya konulması gerekir. Testin neyi ölçmek istediği ortaya konulduktan sonra aynı amaca yönelik alanyazında yer alan başarı testleri ve test maddelerinin yazımında dikkat edilmesi gereken önemli noktalar göz önünde bulundurulmalıdır. Bu sebeple incelenen doktora tezlerinde başarı testleri hazırlanırken amaç belirleme ve alanyazın taraması yapılıp yapılmadığına dikkat edilmiştir.

***Belirtke tablosu****:* Ölçme aracının kapsam geçerliliğini sağlamada kılavuzluk yapan belirtke tabloları madde havuzunda yer alan soruların hangi bilişsel alanlarda olduğunu net bir biçimde göstermektedir. Belirtke tablosunda bilişsel alan sınıflaması, soru maddeleri ve kazanımlar yer almalıdır. Bu sayede tüm maddeler bilişsel alan sınıflaması ve belirtke tablosu dikkate alınarak yazılır ve uzman görüşüne sunulabilir (Gronlund, 1977).

***Madde havuzu (Soru oluşturma):*** Alanyazın taraması yapıldıktan sonra taslak bir başarı testi oluşturma süreci başlamalıdır. Başarı testi hazırlama sürecinin en önemli aşaması soru oluşturmadır. Bu süreçte ölçülecek kazanımlar ile ilgili olarak belirtke tablosu ile birlikte geniş kapsamlı bir madde havuzu hazırlanmalı ve her bir kazanıma yönelik mümkün olduğunca en fazla sayıda madde üretilmeye çalışılmalıdır. Maddelerin hedef kitlenin özelliklerine uygun olarak hazırlanması oldukça önemlidir (Adıgüzel, 2016). Maddeler oluşturulduktan sonra bilişsel alan sınıflamasının da yer aldığı belirtke tablosunda kapsam geçerliği çalışmaları için uzman görüşüne sunulmalıdır (Alpar, 2012).

***Uzman görüşü ve kapsam geçerliği çalışmaları:*** Madde havuzu hazırlandıktan sonra uzmanlar tarafından kapsamı, bilimselliği, dil ve anlatımı ile maddelerin teknik özellikleri bakımından gözden geçirilmelidir. Bu amaçla testi inceleyecek uzmanlar arasında alan uzmanı, dil uzmanı ve ölçme değerlendirme uzmanlarının olması önerilmektedir (Metin, 2014). Uzmanların maddeler ile ilgili görüşlerinin uzman değerlendirme formları kullanılarak alınması bu süreci daha güvenilir kılacaktır. Formlardan elde edilen veriler üzerinde Lawshe tekniği ya da Miles ve Huberman formülü kullanılarak kapsam geçerliği sağlanmalıdır (Lawshe tekniği ile ilgili geniş okumalar Lawshe (1975) kaynağından yapılabilir).

***Pilot uygulama:*** Uzman görüşleri sonrası gerçekleştirilen kapsam geçerliği çalışmaları bazı maddelerin çıkarılması, düzeltilmesi ve yeni maddeler eklenmesi ile tamamlanacaktır. Bu noktadan sonra oluşan taslak başarı testi, esas gruba benzer özellikler taşıyan bir gruba uygulanmalıdır. Gorard (2010) pilot çalışmaları “kıyafet provalarına” benzetir. Bu uygulama bir kez (güvenirlik analizi için iki yarıya bölme yöntemi ya da eş değer formlar yöntemleri kullanılır) yapılabileceği gibi iki kez de (güvenirlik analizleri için test tekrar test yöntemi ya da eş değer formlar yöntemi kullanılır) yapılabilir. Pilot uygulama için seçilecek grup sayısı ile ilgili olarak alanyazında farklı görüşler vardır. Bu görüşler madde sayısının 5 ya da 10 katı sayısınca kişi, en az 50 kişi, en az 100 kişi, en az 200 kişi biçiminde değişmektedir.

***Madde analizleri:*** Madde analizleri kapsamında her bir maddenin madde ayırt edicilik ve madde güçlük indeksleri hesaplanmalı ve raporlanmalıdır. Her bir maddenin madde ayırt edicilik ve madde güçlük indekslerinin alanyazında yer alan değerler arasında olması testin geçerli bir test olduğu anlamına gelmektedir. Madde ayırt edicilik indeksi maddenin yanıtını bilen ile bilemeyen bireyleri ayırma gücünü göstermektedir. İstatistiksel olarak ise toplam test puanları ile madde puanları arasındaki korelasyondur ve indeks -1 ile +1 arasında değişmektedir. Bu değerin negatif olması maddenin bireyleri ters ayırt ettiğinin bir ispatıdır ve negatif değer alan maddeler teste dahil edilmez (Fan ve Sun, 2013). Alanyazında bu değerin 0,30 ve üzerinde olması gerektiği belirtilmiştir (Hinkle, Wiersma ve Jurs, 2003).

Madde güçlük indeksi ise bir maddeye doğru yanıt verenlerin sayısının teste katılan toplam kişi sayısına oranıdır. Örneğin bir maddeye 100 kişilik gruptan sadece 12 kişi doğru cevap vermiş olsun. Bu durumda madde güçlüğü 0,12 olarak bulunur. Bu değerde maddenin çok zor olduğunu gösterir. Başarı testlerinde madde güçlüğü değerinin 0,20 ile 0,80 arasında homojen bir biçimde olması önerilmektedir (Christensen, Johnson ve Turner, 2015). Maddelerin büyük bir kısmı orta düzeyde (0,40-0,60), zor ve kolay sorular ise daha az sayıda olmalıdır.

***Güvenirlik analizleri****:* Başarı testinin geçerlilik ile ilgili işlemleri tamamlandıktan sonra güvenirliğinin test edilmesi gerekir. Salkind (2015) güvenirliği bir testin ya da ölçüm yapmak için kullanılan herhangi bir araca ait ölçümlerin ne kadar tutarlılığı olduğu biçiminde tanımlamıştır. Testin iç tutarlılığı da diyebileceğimiz güvenirlik değerini hesaplamanın birçok yolu vardır. Bunlar Kuder-Richardson 20 ve 21 (KR20 ve KR21), Cronbach alfa katsayısı, Spearman Brown katsayısı ve diğer başka yöntemler biçimindedir (Salkind, 2015).

**Verilerin Analizi**

Araştırmacı, “Başarı Testi Geliştirme Aşamaları Rubriğinde” yer alan 8 kategori açısından 67 doktora tezinde yer alan başarı testlerini incelemiştir. İncelemeler sonucunda 5 doktora tezinde alanyazında var olan başarı testleri kullanıldığı tespit edilmiş olup bu tezler inceleme dışında tutulmuştur. Ayrıca incelen başarı testlerinin tamamı çoktan seçmeli sorulardan oluşan testlerdir. Verilerin analizi 14 üniversitede gerçekleştirilen 62 doktora tezi üzerinden gerçekleştirilmiştir. Araştırmacı, rubrikte bulunan ilgili kategorinin tezde ne düzeyde gerçekleştirilip gerçekleştirilmediğine karar vermiştir. Bu bağlamda kategorilerin gerçekleşme durumları üç düzeyde ele alınmıştır. Bu düzeyler ilgili kategorinin tezde “yeterli düzeyde”, “kısmen yeterli düzeyde” ve “yetersiz düzeyde” yer aldığını ifade etmektedir. Örneğin; tezde uzman görüşü alındığı belirtilmiş ancak uzmanların sayısı, hangi alanda uzman oldukları ve uzmanlardan gelen dönütler belirtilmemişse “Uzman görüşü ve kapsam geçerliği çalışmaları” kategorisi “yetersiz düzey” olarak değerlendirilmiştir. Benzer biçimde “Belirtke tablosu” kategorisi için incelemeler yapıldığında kazanım ve soru maddelerinin yer aldığı tablolar belirtke tablosu olarak ifade edilen tezler de bu kategoride yetersiz düzey olarak değerlendirilmiştir. Araştırmada elde edilen veriler ile ilgili olarak belirlenen kategoriler üzerinden sayısallaştırma gerçekleştirilerek betimsel analiz yapılmıştır. Betimsel analiz yapılabilmesi için yeterli düzey “1”, kısmen yeterli düzey “0,5” ve yetersiz düzey “0” olarak sayısallaştırılmıştır. Buna göre tüm kategorilerin yeterli düzeyde olduğu başarı testi 8 puan alacaktır.

**Güvenirlik**

Doktora tezlerinde yer alan başarı testlerinin tek araştırmacı tarafından incelemesi araştırma bulgularında nesnelliği zedeleyeceğinden ikinci bir puanlayıcı tarafından kura yoluyla belirlenen 20 doktora tezi (%32) puanlanmıştır. Bu puanlama ölçme ve değerlendirme alanında lisanüstü eğitim almış ve bir üniversitede öğretim üyesi olarak görev yapan bir uzman tarafından ”Başarı Testi Geliştirme Aşamaları Rubriği” kullanılarak gerçekleştirilmiştir. Puanlayıcılar arası güvenirlik hesaplaması Miles ve Huberman (2015) tarafından önerilen formül yoluyla gerçekleştirilmiştir.

(Güvenirlik = Görüş Birliği / (Görüş Birliği + Görüş Ayrılığı).

Hesaplama sonucu puanlayıcılar arası uyuşum yüzdesi %84 olarak bulunmuştur. Güvenirlik hesaplarının %70’in üzerinde çıkması, araştırma için güvenilir kabul edilmektedir (Miles ve Huberman, 2015). Buradan elde edilen sonuç araştırma için güvenilir kabul edilmiştir.

**BULGULAR**

**Başarı Testi Geliştirme Aşamalarına İlişkin Bulgular**

Başarı testi geliştirme süreci bütüncül bir süreçtir. Her bir aşama bir sonraki aşamanın öncülü olmak zorundadır. Bu bağlamda ilk olarak tüm aşamaların gerçekleşme durumlarına bakılmış ve hazırlanan başarı testlerinin bütüncül olarak hangi düzeyde yeterli olduğu incelenmiştir. Bu inceleme toplam 8 kategoriyi sayısallaştırmak için verilen “1”, “0,5” ve “0” puanları toplanarak yapılmıştır. Tüm doktora tezlerinde yer alan başarı testlerine ilişkin puanlamalar Tablo 2 de gösterilmiştir.

Tablo 2.

Tüm başarı testlerinin aldıkları puanlara ilişkin bilgiler

|  |  |  |
| --- | --- | --- |
| Puan | Tez sayısı (n) | Yüzde (%) |
| 2,5 | 4 | 6,5 |
| 3 | 4 | 6,5 |
| 3,5 | 3 | 4,8 |
| 4 | 3 | 4,8 |
| 4,5 | 5 | 8,1 |
| 5 | 7 | 11,3 |
| 5,5 | 9 | 14,5 |
| 6 | 13 | 21,0 |
| 6,5 | 3 | 4,8 |
| 7 | 8 | 12,9 |
| 8 | 3 | 4,8 |
| Toplam | 62 | 100,0 |

Tablo 2’de görüldüğü gibi doktora tezlerinde yer alan başarı testlerinin araştırmacı tarafından geliştirilen rubrikten aldığı puanlar 2,5 ile 8 arasında değişmektedir. En yüksek puan olan 8, başarı testi geliştirme aşamalarının tamamımın yeterli düzeyde gerçekleştirildiği anlamına gelmektedir. Bu bağlamda tüm aşamaların yeterli düzeyde gerçekleştirildiği başarı testi sayısı sadece 3 (%4,8) tür. Kapsam geçerliği çalışmaları dışında tüm kategorilerin yeterli düzeyde gerçekleştirildiği başarı testi sayısı ise 8’dir (%12,9). Başarı testlerinin nitelik düzeyinin toplam puan ortalaması 5,3 olarak hesaplanmıştır. Bu bulgu incelenen başarı testlerinin genel olarak orta düzey niteliğe sahip olduğunu göstermektedir. Tablo 3 de ise her bir kategorinin ortalama puanları (0-1 arasında) sunulmuştur.

Tablo 3.

Başarı testleri geliştirme kategorilerine ait ortalama puanlar

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Kategori | Amaç bel. alanyazın incelemesi | Belirtke tablosu | Madde havuzu | Uzman görüşü | Kapsam geçerliği | Pilot uygulama | Madde analizleri | Güvenirlik çalışmaları |
| x̄ | 0,74 | 0,33 | 0,85 | 0,65 | 0,11 | 0,95 | 0,71 | 0,94 |
| ss | 0,41 | 0,47 | 0,34 | 0,37 | 0,32 | 0,21 | 0,43 | 0,20 |

Tablo 3’e bakıldığında en düşük ortalama puanın, kapsam geçerliği çalışmaları (x̄: 0,11) ve belirtke tablosu (x̄: 0,33) kategorilerinde olduğu görülmektedir. Bu kategoriler yetersiz olarak değerlendirilebilir. Pilot uygulama (x̄: 0,95) ve güvenirlik (x̄: 0,94) kategorileri ise yeterli düzeyde en yüksek ortalama puanlara sahiptir. Başarı testi geliştirme aşamalarına ait diğer bulgular üç ana başlık altında sunulacaktır. Bunlar; hazırlık, uygulama ve analiz başlıklarıdır.

***Hazırlık Aşamasına İlişkin Bulgular***

Hazırlık başlığında amaç belirleme ve alanyazın taraması, madde havuzu (soru oluşturma) ve belirtke tablosu aşamaları yer almaktadır. Bu aşamalar ile ilgili bulgular Grafik 1’de sunulmuştur.

Grafik 1.

Hazırlık aşamasına ait bulgular

Araştırma kapsamında incelenen 62 doktora tezinde yer alan başarı testlerinin 43’ünde (%69,4) amaç belirleme ve alanyazın taramasının yeterli düzeyde yapıldığı, 6’sında (%9,7) kısmen yeterli olduğu ve 13’ünde (%21) amaç belirleme ve alanyazın taraması ile ilgili ifadelerin yer almadığı tespit edilmiştir. İncelenen doktora tezlerinin 52’sinde (%83,9) madde havuzu (soru oluşturma) hazırlama süreci detaylı bir biçimde açıklanmışken 8’inde (%12,9) madde havuzu (soru oluşturma) hazırlama sürecinden bahsedilmemiştir. 2 tezde (%3,22) ise madde havuzu (soru oluşturma) hazırlama süreci kısmen yeterli bulunmuştur. Hazırlık başlığının son aşaması olan belirtke tablosu, tezlerin 41’inde (%66,1) yetersiz bulunmuştur. Bu yetersizliğin bir kısmı belirtke tablosuna hiç yer verilmemesinden, bir kısmı da belirtke tablosunda bilişsel alan sınıflamalarına, kazanımlara ya da madde numaralarına yer verilmemesinden kaynaklanmaktadır. İncelenen tezlerin 20’sinde (%32,3) tam anlamıyla belirtke tabloları yer almıştır. 1 (%1,6) tezde ise belirtke tablosu kazanımlara göre yapılmamış olup, konular ve madde numaralarına göre hazırlanmıştır. Bu bulgular hazırlık başlığında özellikle belirtke tablolarının hazırlanması konusunda ciddi sorunlar olduğunu göstermektedir.

***Uygulama Aşamasına İlişkin Bulgular***

Uygulama başlığında uzman görüşü, kapsam geçerliği çalışmaları ve pilot uygulama aşamaları yer almaktadır. Bu aşamalar ile ilgili bulgular Grafik 2’de sunulmuştur.

Grafik 2.

Uygulama aşamasına ait bulgular

Araştırmada incelenen 62 doktora tezinde yer alan başarı testlerinin 29’unda (%46,8) uzman görüşü aşamasının yeterli düzeyde yapıldığı söylenebilir. Bu 29 tezde uzmanların kim olduğu, hangi ölçütlerle belirlendiği, görüşleri sonrasında hangi düzeltmelerin yapıldığı detaylı olarak belirtilmiştir. Toplam 10 (%16,1) tezde uzman görüşü ile herhangi bir bilgi verilmezken 23 (%37,1) tezde ise sadece uzman görüşüne başvurulduğu ifade edilmiştir. Uzman görüşleri sonrası yapılması gereken kapsam geçerliği çalışmaları sadece 7 (%11,3) tezde Miles ve Huberman formülü kullanılarak yapılmıştır. Geri kalan 55 (%88,7) tezde ise uzman görüşü sonrası kapsam geçerliği çalışmalarına rastlanılmamıştır. Son olarak tezlerde yer alan başarı testlerinin 59’una (%95,2) geçerlik ve güvenirlik çalışmaları için pilot uygulama yapılmış, 3’üne (%4,8) ise yapılmamıştır. Bu bulgular, uygulama başlığında özellikle uzman görüşü ve uzman görüşü sonrası kapsam geçerliği çalışmaları konusunda önemli problemler olduğunu yansıtmaktadır.

***Analiz Aşamasına İlişkin Bulgular***

Analiz başlığında madde analizleri (madde ayırtedicilik ve madde güçlüğü) ve güvenirlik analizlerinin sunulması aşamaları yer almaktadır. Bu aşamalar ile ilgili bulgular Grafik 3’te sunulmuştur.

Grafik 3.

Analiz aşamasına ait bulgular

İncelenen başarı testlerinin 42’sinde (%67,7) madde analizleri yeterli düzeyde yapılmışken, 15 (%24,2) başarı testine ait madde analizleri yetersiz bulunmuştur. Başarı testlerinin 57’sinde (%91,9) detaylı açıklamalar ile birlikte güvenirlik katsayısına yer verilmiş, 3’ünde (%4,8) ise güvenirlik katsayısına yer verilmiş ancak katsayı türü açıklanmamıştır. Bu bulgular başarı testi geliştirme sürecinin analiz aşamasında eksiklikler olmasına rağmen çoğunlukla bu aşamaya uygun işlemler yapıldığını göstermektedir.

Araştırma kapsamında incelenen 62 başarı testinin madde analizlerinin gerçekleştirilmesinde SPSS, TAP, FINESSE, ITEMAN, EXCEL gibi çeşitli paket programlar kullanılmıştır. Güvenirlik katsayılarının hesaplanmasında da test tekrar test ve iki yarıya bölme yöntemine başvurulmuş ve katsayılar Cronbach alfa, KR 20, Spearman Brown, Pearson momentler çarpımı ve korelasyon olarak ifade edilmiştir. Madde analizleri kapsamında sadece 1 tezde açımlayıcı ve doğrulayıcı faktör analizi gerçekleştirilmiştir. Ancak başarı testlerinde olduğu gibi puanlamanın kategorik yapıldığı durumlarda açımlayıcı faktör analizinin tetrakorik korelasyon matrisi üzerinden yapılması ve açımlayıcı faktör analizi varsayımlarının sağlanması gerektiği düşünüldüğünde (Şekercioğlu, Bayat ve Bakır, 2014) bu tezde yer alan başarı testi ile ilgili analizlerin yetersiz olduğu söylenebilir.

**TARTIŞMA, SONUÇ VE ÖNERİLER**

Bu araştırma ile Türkiye’de ilköğretim anabilim dalı bünyesinde 2012-2017 yılları arasında gerçekleştirilen doktora tezlerinde ölçme aracı olarak kullanılan başarı testlerinin geliştirilme süreçleri incelenmiştir.

Araştırma kapsamında toplam 62 doktora tezinde ölçme aracı olarak kullanılan başarı testleri incelenmiştir. İnceleme sonucunda başarı testi geliştirmenin her aşamasında yetersizlikler olduğu sonucuna ulaşılmıştır. Bu yetersizlikler en çok hazırlık aşamasında belirtke tablolarının hazırlanmasında (%66,1), uygulama aşamasında uzman görüşü alma (%53,2) ve kapsam geçerliği çalışmalarında (%88,7) görülmektedir. Bu sonuçlar alanyazında yer alan benzer araştırma sonuçlarıyla örtüşmektedir (Hall, 1986; Karadağ, 2011; Mutluer ve Yandı, 2012; Onwuegbuzie, 2002). Başarı testi geliştirme sürecinde kapsam geçerliği açısından birbirini tamamlayan aşamalar olduğundan uzman görüşüne sunulan taslak başarı testi belirkte tablosu ile birlikte sunulmalıdır. Uzmanlardan gelen düzeltmeler kapsam geçerliği çalışmalarında kullanılarak test, pilot uygulama için hazır hale getirilecektir. Bu işlemler ciddi emek ve dikkat gerektirir. Ayrıca başarı testi geliştirme süreçlerinin hazırlık ve uygulama aşamalarına olan benzerliği dikkate alındığında ölçek geliştirme süreçlerini inceleyen araştırmalarda da (Çüm ve Koç, 2013; Delice ve Ergene, 2015; Erkuş, 2000; Erkuş, 2007; Erkuş, 2016; Karadağ, 2011; Mor Dirlik, 2014; Şahin ve Boztunç Öztürk, 2018) ölçeklerin nitelik düzeyleri yetersiz olarak tespit edilmiştir. Bu yetersizlikler daha çok geçerlik ile ilgili özelliklerde görülmektedir.

Başarı testi geliştirme sürecinin son basamağı olan analiz aşamasında ise eksikliklerin diğer aşamalara göre az olması dikkati çeken bir bulgudur. Bu bulgu araştırmacılar tarafından başarı testi geliştirme sürecinde analiz aşamasına verilen önemin hazırlık ve uygulama aşamalarına verilen önemden daha fazla olduğunu gösterir niteliktedir. Karadağ (2011) tarafından yapılan araştırmada ise geçerlik ve güvenirlik saptamalarında yetersizlikler olduğu sonucuna ulaşılmıştır. Başarı testlerinin geliştirilmesinde kapsam geçerliği ile ilgili eksiklikler devam ederken, madde analizleri ve güvenirlik ile ilgili bilgilere yer verilmesi son yıllarda başarı testi geliştirme konusunda olumlu ilerlemelerin olduğunu göstermektedir. Bu sonuç üzerinde son yıllarda eğitim bilimleri alanında verilen istatistik dersleri ve yayımlanan istatistik kitaplarında ki artışın etkisinin olabileceği düşünülmektedir. Elbette bu çıkarımı daha net ve veriye dayalı olarak yapabilmek için madde analizleri üzerine odaklanan yeni araştırmalar tasarlamak yerinde olacaktır.

Başarı testi geliştirme işi bütüncül bir süreç gerektirir. Bu süreçte yapılan her işlem detaylı bir biçimde açıklanmalıdır. Her bir aşamanın yeterli düzeyde olduğu başarı testi sayısının sadece 3 (%4,8) olması tezlerde kullanılan başarı testlerinin istenilen niteliğe sahip olmadığına işaret etmektedir. Bu sonuç, eğitim araştırmalarının nitelik sorgulamaları kapsamında yapılan diğer araştırma sonuçları (Boyraz ve Serin, 2016; Evrekli, İnel, Deniş ve Balım, 2011; Kabaca ve Erdoğan, 2007; Karadağ, 2010; Konan ve Oğuz, 2015; Şahin, Calp, Bulut ve Kuşdemir, 2013) ile örtüşmektedir.

Bu araştırma, ilköğretim anabilim dallarında gerçekleştirilen doktora tezlerinde ölçme aracı olarak geliştirilen başarı testlerinin nitelik düzeylerini net bir biçimde ortaya koymaya çalışmıştır. Elbette incelenen başarı testlerinin tek bir araştırmacı tarafından değerlendirilmesi araştırmanın en önemli sınırlılıklarından birisidir. Ancak değerlendirmelerde yapılabilecek hata oranı, başarı testlerinin nitelik anlamında yetersizliğinin savunucusu olmaya yetmeyecektir.

Araştırmadan elde edilen bulgular düşünüldüğünde şu öneriler yapılabilir:

* Mevcut İlköğretim anabilim dalı doktora programlarında mutlaka ölçme araçlarının geliştirilmesi içeriğine sahip dersler konulmalı ve bu dersler alanın uzmanları tarafından verilmelidir.
* Doktora tez danışmanları, tez izleme komitesi üyeleri ve tez jüri üyeleri ölçme aracı olarak kullanılan başarı testlerinin niteliği konusunda gerekli dikkati göstermeye özen göstermelidir.
* Eğitim alanında gerçekleştirilen doktora tezlerinin eksikliklerini ortaya koyması ve böylece gelecekte yapılacak olan doktora tezlerinin kalitesinin artmasına katkı sağlayacağı düşünüldüğünde ölçme araçlarına yönelik nitelik sorgulama araştırmalarının periyodik olarak gerçekleştirilmesi gerekmektedir.
* Bu gibi araştırmaların kapsamlı bir araştırmacı topluluğu tarafından yapılması daha geniş ve daha net sonuçlara ulaşmakta kolaylık sağlayacaktır.

**KAYNAKÇA/REFERENCES**

Adıgüzel, O. C. (2016). *Eğitim programlarının geliştirilmesinde ihtiyaç analizi el kitabı.*Ankara: Anı Yayıncılık.

Alpar, R. (2012*). Uygulamalı istatistik ve geçerlik-güvenirlik*. Ankara: Detay Yayıncılık

Anastasi, A. (1988). *Psychological testing*. NewYork: Macmillan Publishing Company.

Bayık, M. E. & Gürbüz, S. (2016). Ölçek uyarlamada metodoloji sorunu: Yönetim ve örgüt alanında uyarlanan ölçekler üzerinden bir araştırma*. İş ve İnsan Dergisi, 2016; 3(1), 1-20*.

Boyraz, C. & Serin G. (2016, Mayıs). *Sınıf öğretmenliği alanında yapılan lisansüstü tezlerin (2011 2015) deneysel desen ölçütlerini karşılama durumu*. XVIII. AMSE-AMCE-WAER Kongresinde sunulan sözlü bildiri. Anadolu Üniversitesi, Eskişehir.

Büyüköztürk, Ş., Çakmak, E. K., Akgün, O. E., Karadeniz, Ş., & Demirel, F. (2014). *Bilimsel araştırma yöntemleri.* Ankara: Pegem Akademi

Christensen, L. B., Johnson, R. B., & Turner, L. A. (2015). *Araştırma yöntemleri desen ve analiz*. A. Aypay (Çeviri Edt.). Ankara: Anı Yayıncılık

Çüm, S. & Koç, N. (2013). Türkiye’de psikoloji ve eğitim bilimleri dergilerinde yayımlanan ölçek geliştirme ve uyarlama çalışmalarının incelenmesi*. Eğitim Bilimleri ve Uygulama, 12 (24), 115-135.*

Delice, A. & Ergene, Ö. (2015). Ölçek geliştirme ve uyarlama çalışmalarının incelenmesi: Matematik eğitimi makaleleri örneği. *Karaelmas Journal of Educational Sciences, 3, 60-75.*

Erkuş, A. (2000). Sık kullanılan bazı psikolojik ölçeklerin güvenirliklerinin irdelenmesi. *Türk Psikoloji Yazıları, 3(6), 3-17.*

Erkuş, A. (2007). Ölçek geliştirme ve uyarlama çalışmalarında karşılaşılan sorunlar. *Türk Psikoloji Bülteni, 13(40) 17.*

Erkuş, A. (2016). Ölçek geliştirme ve uyarlama çalışmalarındaki sorunlar ile yazım ve değerlendirilmesi. *Pegem Atıf İndeksi, 1221-1234. doi:10.14527/9786053183563.075*

Evrekli, E., İnel, D., Deniş, H., & Balım, A. G. (2011). Fen eğitimi alanındaki lisansüstü tezlerdeki yöntemsel ve istatistiksel sorunlar. *İlköğretim Online*, *10*(1).

Fan, X. & Sun, S. (2013). Item response theory. T. Teo (Edt.) In *Handbook of quantitative methods for educational research* Sense Publishers.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education.* New York: McGram-Hill Companies.

Gorard, S. (2010). Q*uantitative methods in educational research: the role of numbers made easy*. Bloomsbury Publishing.

Gronlund, N. E. (1977). *Constructing achievement test*. New Jersey: Prentice Hall, Inc.

Hall, B. W. (1986). Validity, reliability, and norms of popular versus less popular published educational achievement tests. *Journal of Educational Research, 79 (3), 145-150*

Hinkle, D.E., Wiersma, W. & Jurs, S.G. (2003). *Applied statistic for the behavioral sciences.* Boston: Houghton Mifflin Company.

Kabaca, T. & Erdoğan, Y. (2007). Fen bilimleri ve matematik eğitimi alanlarındaki tez çalışmalarının istatistiksel açıdan incelenmesi. *Pamukkale Eğitim Fakültesi Dergisi, 2(22), 54-63.*

Karadağ, E. (2009). Eğitim bilimleri alanında yapılmış doktora tezlerinin tematik açıdan incelemesi. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, *10*(3).

Karadağ, E. (2010). Eğitim bilimleri doktora tezlerinde kullanılan araştırma modelleri: Nitelik düzeyleri ve analitik hata tipleri. *Kuram ve Uygulamada Eğitim Yönetimi*, *1*(1), 49-71.

Karadağ, E. (2011). Eğitim bilimleri doktora tezlerinde kullanılan ölçme araçları: Nitelik düzeyleri ve analitik hata tipleri*. Kuram ve Uygulamada Eğitim Bilimleri, 11(1), 311-334.*

Konan, N. & Oğuz, V. (2015). Yüksek lisans ve doktora tezlerinde denence kullanımı. *Electronic International Journal of Education, Arts, and Science (EIJEAS)*, *1*(2).

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28, 563–575.*

McMillan, J.H. (2004). *Educational research: Fundamentals for the consumer*. Boston: Pearson Education, Inc

Metin, M. (2014). Nicel veri toplama araçları. M. Metin (Edt.) *Kuramdan uygulamaya eğitimde bilimsel araştırma yöntemleri* içinde, Ankara: Pegem Akademi

Miles, M. B. & Huberman, A. M. (2015). Nitel veri analizi (İkinci Baskıdan Çeviri). (Çev: S. Akbaba-Altun ve A. Ersoy). Ankara: Pegem A Akademi Yayınları.

Mor Dirlik, E. (2014). Ölçek geliştirme konulu doktora tezlerinin test ve ölçek geliştirme standartlarına uygunluğunun incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 5 (2), 62-78*

Mutluer, C. & Yandı, A. (2012, Eylül). *Türkiye’deki üniversitelerde 2010-2012 yılları arasında yayımlanan tezlerdeki başarı testlerinin incelenmes*i. III Ulusal Eğitim ve Psikolojide Ölçme ve Değerlendirme Kongresinde sunulan sözlü bildiri. Abant İzzet Baysal Üniversitesi, Bolu.

Onwuegbuzie, A. J. (2002). Common analytical and interpretational errors in educational research: an analysis of the 1998 volume of the British Journal of Educational Psychology*. Educational Research Quarterly, 26, 11-22.*

Öztürk, N. G., Eroğlu, M. G. ve Kelecioğlu, H. (2015). Eğitim alanında yapılan ölçek uyarlama makalelerinin incelenmesi. *Eğitim ve Bilim, 40(178), 123-137*

Salkind, N.J. (2015). *İstatistikten nefret edenler için istatistik*. A. Çuhadaroğlu (Çev. Edt.) Ankara: Anı Yayıncılık

Şahin, D., Calp, Ş., Bulut, P., & Kuşdemir, Y. (2013). Sınıf öğretmenliği eğitimi bilim dalında yapılmış lisansüstü tezlerin çeşitli kriterlere göre incelenmesi. *Zeitschrift für die Welt der Türken/Journal of World of Turks*, *5*(3), 187-205.

Şahin, M , Boztunç Öztürk, N . (2018). Eğitim Alanında Ölçek Geliştirme Süreci: Bir İçerik Analizi Çalışması. *Kastamonu Eğitim Dergisi, 26 (1), 191-199. DOI: 10.24106/kefdergi.375863*

Şekercioğlu, G., Bayat, N., & Bakır, S. (2014). Fen maddelerini anlama testinin psikometrik niteliklerinin belirlenmesi. *Eğitim ve Bilim*, *39*(176).

Tan, Ş. (2006). *Öğretimde planlama ve değerlendirme.* Ankara: Pegem Akademi.

Turgut, M. F. & Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi.

**İletişim/Correspondence**

Arş. Gör. Celal BOYRAZ

cboyraz@bayburt.edu.tr