

Ölçek Geliştirme Konulu Doktora Tezlerinin Test ve Ölçek Geliştirme Standartlarına Uygunluğunun İncelenmesi

The Analysis of the Doctoral Dissertations Themed of Scale Development According to the Test and Scale Development Standards

Ezgi Mor Dirlik*

Öz

Psikolojik testler eğitim bilimlerinden, sağlık bilimine, gazetecilikten turizme kadar çok geniş bir yelpazede sıklıkla kullanılan ölçme araçlarıdır. Özellikle bilimsel araştırmalarda oldukça sık tercih edilen bu ölçme araçları, gerek uyarılma gerekse ölçek geliştirme çalışmalarıyla ilgili alanlara kazandırılmakta ve kullanıcılara sunulmaktadır. Geliştirilen psikolojik testlerin amacına yönelik ölçmeler yapabilmesi yeni çalışmalarla takip edilmesi ve niteliklerinin sorgulanması gerekmektedir. Bu çalışma kapsamında, APA tarafından yayınlanan Eğitimde ve Psikolojide Ölçme Standartları kitabının test ve ölçek geliştirme bölümünde yer alan standartlardan oluşturulan bir kontrol listesi kullanılarak eğitim bilimleri alanında son beş yıl içerisinde gerçekleştirilen ölçek geliştirme amaçlı beş doktora tezi incelemeye alınmıştır. İncelemeye alınan doktora tezlerinin, söz konusu standartları karşılama durumu kategorisel içerik analizi yöntemi kullanılarak belirlenmiştir. Her psikolojik test için karşılanan, kısmen karşılanan ve hiç karşılanmayan standartların frekansları belirlenmiş, böylece söz konusu doktora tezlerinde uyulması gereken ölçek geliştirme standartlarının ne düzeyde karşılandığı belirlenmiştir. Sonuçlar incelendiğinde, incelenen tezlerden ikisinin standartları karşılama düzeyinin oldukça yüksek olduğu, bir tez çalışmasının standartları orta düzeyde ve diğer iki tez çalışması ise incelenen standartları oldukça düşük düzeyde karşıladığı belirlenmiştir. Her tez çalışmasında geliştirilen psikolojik testlerin gözden geçirilmesi gereken yönleri de yapılan incelemeler sonucunda belirlenmiştir.

Anahtar Kelimeler: Ölçek Geliştirme Tezleri, APA Test ve Ölçek Geliştirme Standartları..

Abstract

Psychological tests are used so widely from educational sciences to journalism and tourism. These measurement tools which are preferred especially in scientific research are prepared both adaption and development studies and presented to the users. In order to measure the psychological constructs according to the goal of development, these tests should be revised and controlled with new studies. In the context of this study, five doctoral dissertations published in the last five years with the aim of development of scale in educational sciences were selected randomly and analyzed by using the checklist which was prepared based on the Education and Psychological Testing Standards book published by American Psychological Association in 1999. The situations of these dissertations according to the standards were determined with the categorical content analysis. For each dissertation, the frequencies of the standard which are totally agreed, partial agreed and absent were calculated, so the ratio of these standards which were taken into account in scale development was defined. As a result of this study, it has been found that two of dissertations met the test and scale development standards so highly, one of dissertations met these standards moderately and the last two dissertations met these standards in low level. Also the points which should be revised in these dissertations according to the standards were defined.

Key Words: The Dissertations of Scale Development, APA, Test and Scale Development Standards

* Arş. Gör., Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Ankara-Türkiye, e-posta: edirlik@ankara.edu.tr

GİRİŞ

Modern hayatta akla gelebilecek her türlü alanda bireylere daha iyi hizmet sunmak adına, bireylere ilişkin çok çeşitli bilgiler toplanmaktadır. Çeşitli alanlarda hizmet veren kurumlar tarafından müşterilerin memnuniyetleri ve ürün tercihleri incelenmekteyken, işe yerleştirme, eleman seçme gibi alanlarda da bireylerin kişisel ve mesleki özellikleri incelenmektedir. Bireyler hakkında bilgi toplamanın en yaygın olduğu alanların başında şüphesiz eğitim gelir. Temel eğitimden yükseköğrenime kadar, eğitimin her aşamasında eğitimciler tarafından öğrencilerin akademik yeterlikleri, öğrenme düzeyleri ve duyuşsal özellikleri ölçülmektedir. Her alanda gerçekleşen bu ölçme faaliyetlerinde en sık kullanılan yöntem bireyi bir test ile karşı karşıya getirmek ve bireyin testteki maddelere gösterdiği tepkileri belli bir yöntemle göre puanlamaktır (Anastasi, 1988), yani psikolojik testlerden yararlanmaktır. Sıklıkla tercih edilen psikolojik testler, gözlem, görüşme gibi diğer bilgi toplama yöntemlerine göre daha kolay uygulanabilmekte, objektif olarak puanlanabilmekte ve geçerli ve güvenilir gözlemler yapmaya olanak sağlamaktadır (Conway, 2006; Cronbach, 1990; Pawar ve Pawar, 2008).

Bireyler hakkında bilgi edinmek için sıklıkla tercih edilen psikolojik testlere dair alan yazında birden fazla tanım yapılmıştır. Anastasi (1988), test sözcüğünü belirli bir davranışın nesnel ve standart koşullarda ölçülmesi olarak tanımlamıştır. Cronbach (1990)'a göre psikolojik testler, iki veya daha fazla bireyin davranışını sistematik olarak karşılaştırma amacıyla kullanılan araçlardır. Urbina (2004)' ya göre ise psikolojik ölçme araçları, insanlar ve onların davranışları hakkında çıkarım yapma konusunda araştırmacılara yardımcı olan araçlardır. Cansever (1982)' e göre psikolojik test, kişinin davranışını standart koşullar altında gözlemlemek ve tanımlamak için kullanılan sistematik yöntemdir. Özoğlu (2007), psikolojik testleri, bireylerin psikolojik özelliklerini ölçmeyi amaçlayan araçlar olarak tanımlamıştır. Yapılan tüm tanımlardan da anlaşıldığı üzere, psikolojik testler bireylerin davranışlarını gözlemleme, gözlemlenen davranışlardan çıkarımlar yapma ve birey davranışlarını karşılaştırma amacıyla kullanılmaktadır. Bu kullanımlara ek olarak, psikolojik testler bilimsel çalışmalarda da bilgi toplama amacıyla sıklıkla kullanılmakta, özellikle de sosyal bilimler alınan yapılan yüksek lisans ve doktora tezlerinde ölçek geliştirme çalışmalarına rastlanmaktadır. Yüksek Öğretim Kurulu'na bağlı olarak faaliyet gösteren Ulusal Tez Merkezi veri tabanında "ölçek geliştirme" anahtar sözcüğü kullanılarak tarama yapıldığında, hem doktora hem de yüksek lisans düzeyinde farklı bilimsel alanlarda yapılmış 172 çalışma bulunmaktadır ve bu çalışmaların yaklaşık %40'ı eğitim bilimlerinde gerçekleştirilmiştir. Geriye kalan çalışmalar ise, tıp, turizm, hemşirelik, işletme, biyoloji, halk sağlığı, spor bilimleri, gazetecilik ve mühendislik gibi farklı alanlarda yapılan ölçek geliştirme çalışmalarıdır (<https://tez.yok.gov.tr/UlusalTezMerkezi/>, 2014). Belirlenen bu değerlerden de anlaşıldığı üzere, araştırmacılar farklı alanlarda çalışıyor olsalar dahi, çalışmalarında ölçme araçlarına ihtiyaç duymakta ve çalışmalarının amacına uygun olarak daha önce geliştirilmiş ya da uyarlanmış uygun bir ölçme aracı bulamazlarsa, ölçek geliştirme faaliyetlerine yönelmektedirler.

Oldukça farklı alanlarda gerçekleştirilen araştırmalarda, bu denli sıklıkla kullanılan ölçme araçları, ancak sistematik bir şekilde standartlara uyularak geliştirilirse ve gereken izleme çalışmaları gerçekleştirilirse, bu ölçeklerden elde edilen bilgiler anlamlı olur (Özoğlu, 2007). Bir başka deyişle psikolojik testlere dayanılarak elde edilen bilgilerin anlamlı olma olasılığını artırmak ve bu ölçme araçlarının geliştirilme amacına daha iyi hizmet etmesini sağlamak için, psikolojik testler büyük bir titizlikle geliştirilmeli, ölçek

geliştirme aşamaları sistematik olarak takip edilmeli ve gereken izleme çalışmaları gerçekleştirilmelidir (Baykul ve Turgut, 2010; Cronbach, 1990).

Gerek bilimsel araştırma amacıyla gerekse ihtiyaçtan kaynaklanan nedenlerle, ülke çapında geliştirilen ölçek sayısı yıllara göre artış göstermiş olsa da, geliştirilen ve uyarlanan ölçeklerle ilgili yapılan araştırmaların oldukça az sayıda olduğu bilinmektedir (Çukur, 1999). Söz konusu ölçekler araştırmacılar tarafından geliştirilmekte, ölçümlere ilişkin geçerlik ve güvenilirlik çalışmaları yapılmakta, fakat daha sonraki yıllar içerisinde çoğu ölçek için izleme çalışmaları yapılmamakta ve başka araştırmacılar tarafından bu ölçme araçlarının psikometrik nitelikleri incelenmemektedir (Koç, 1986; 2008; Öner, 1994; Uçman, 1980). Ülke çapında kullanılmakta olan psikolojik testlere ilişkin yapılan çalışmaların, oldukça sınırlı sayıda olması da bu durumun kanıtıdır. Bu konuda yapılan çalışmalar kronolojik olarak sıralanacak olursa, ilk çalışmanın 1971 yılında Milli Eğitim Bakanlığı (MEB) Planlama Araştırma ve Koordinasyon Dairesi Rehberlik ve Psikolojik Danışma Bölümü'nce gerçekleştirilen bir tarama çalışması olduğu görülmektedir. Bu çalışma kapsamında o yıllarda ülke çapında mevcut olan psikolojik testler taranmış, sınıflandırılmış ve psikometrik nitelikleri bakımından incelenmiştir. Toplamda 20 psikolojik test incelenmiş ve bu testlerin tümünün ABD'de geliştirilmiş olan testler olduğu, Türkiye'de ise doğrudan çeviri ya da uyarlanmış olarak kullanılmakta olduğu ve hiçbirisi üzerinde norm, geçerlik ve güvenilirlik çalışmalarının yapılmadığı belirtilmiştir (MEB; 1971). Özge (1981) yüksek lisans tezi olarak hazırladığı çalışmada, ülke çapındaki psikolojik testleri psikometrik nitelikleri bakımından betimlemeyi (testlerin betimlenmesi ifadesi uygun değil bence) amaçlamış ve 126 psikolojik testi çeşitli nitelikler gözeterek incelenmiştir. Bu çalışmada Türkiye'de o tarihlerde kullanılan 126 psikolojik testin 121'inin yabancı ülkelerde geliştirilen psikolojik testlerden uyarlanmış olduğu, yalnızca beş psikolojik testin Türkiye'de geliştirilmiş olduğu, uyarlama testlerin ise, 84'ünün yalnızca bir çeviri çalışması niteliğinde olduğu belirlenmiştir. Ayrıca tüm psikolojik testler içinde bu testlerle yapılan ölçümlerin geçerlik ve güvenilirliklerine ilişkin kanıt toplayan çalışmalarının her ikisinin de yapıldığı psikolojik test sayısının beş olduğu ve sadece altı testin yayınlanmış el kitabı olduğu sonuçlarına ulaşılmıştır.

Ramazan (1988), Özge (1981) tarafından yapılan çalışmaya benzer bir çalışma yaparak mevcut psikolojik testleri taramış ve Türkiye'de o tarihte kullanılmakta olan psikolojik test sayısını 164 olarak belirlemiştir. Söz konusu psikolojik testlerin 36'sının Türkiye'de geliştirilmiş özgün test, 60 tanesinin uyarlama ve 68'inin ise sadece Türkçe 'ye çevrilmiş olduğu bu çalışmada saptanmıştır. Öngel ve Smith (1994) tarafından gerçekleştirilen çalışmada ise Türk Psikoloji Dergisi'nde 1978-1992 yılları arasında yayınlanan 152 makale incelenmiş ve bu makalelerde kullanılan psikolojik testler çeşitli nitelikleri bakımından incelenmiştir. Bu incelemeye göre, Türkiye'deki araştırmalarda kullanılan psikolojik testlerin % 85'inin Amerika'dan veya Avrupa'dan doğrudan Türkçeye çevrilerek kullanıldığı belirlenmiştir. Çeviri olarak kullanılan testlerin % 5'inin çok az (kelime değişikliği) değişiklik yapılarak veya hiç değişiklik yapılmadan doğrudan Türkçe 'ye çevirisi yapılırken, % 12'sinin yerel kültüre uyarlandığı ve yalnızca % 3'ünün yerel kültüre dayanarak geliştirildiği belirlenmiştir (akt. Çukur, 1999). Psikolojik testler üzerine son yıllarda yapılan çalışmalardan biri ise Doğan tarafından 2009 yılında gerçekleştirilmiştir. Doğan (2009) çalışmada, Türkiye'deki psikoloji alanında kullanılan geliştirilmiş ya da uyarlanmış psikolojik testlerin niceliğini ve niteliğini sistematik bir şekilde belirlemeyi amaçlamıştır. Bu amaçla psikoloji alanında önemli yer tutan dergiler

olan Türk Psikoloji Yazıları (1998-2007), Türk Psikoloji Dergisi (1990-2007) ve Türk Psikiyatri Dergisinde (1990-2007) belirtilen yıllar arasında yayınlanan tüm makaleler bu çalışma çerçevesinde geliştirilen ölçütler bağlamında taranmış ve ilgili betimsel istatistikler rapor edilmiştir. Belirtilen üç dergide toplam 958 makale taranmış ve 389 farklı ölçeğin kullanıldığı belirlenmiştir. Bu ölçeklerin %75'inin uyarlama, %13'ünün özgün ve %12'sinin ise ölçek çevirisi olduğu ve uyarlama çalışmalarının büyük oranda dilsel veya kültürel eşdeğerlik) konusuyla ilgili ölçütleri sağlamadığı belirlenmiştir.

Bu alanda gerçekleştirilen yakın zamanlı bir araştırma ise, araştırmacı tarafından yazılan yüksek lisans tezidir (Dirlik, 2013). Bu çalışmada eğitim kurumlarında sıklıkla kullanılan psikolojik testlerden dördü uzman görüşleri temel alınarak belirlenmiş ve American Psychological Association (APA) tarafından Eğitimde ve Psikolojide Ölçme Standartları olarak belirlenen standartlardan geçerlik, güvenirlik, test geliştirme ve revizyon standartları gözetilerek incelenmiştir. İncelemeye alınan testlerden uyarlanmış olan Temel Kabiliyetler Testi 7-11 yaş formunun uyarlama standartlarını oldukça düşük düzeyde karşıladığı bulunurken, bir diğer uyarlanmış test olan Wechsler Çocuklar İçin Zekâ Ölçeği için ise bu standartların büyük çoğunluğu karşılanmıştır. Türkiye'de geliştirilen psikolojik testlerden Gazi Erken Çocukluk Gelişimi Değerlendirme Aracı ve Akademik Benlik Kavramı Ölçeği ise geçerlik, güvenirlik ve test geliştirme standartları bakımından incelenmiş, özellikle geçerlik ve güvenirlik standartlarının her iki testte de çok düşük düzeyde karşılandığı sonucuna ulaşılmıştır. Test geliştirme ve uyarlama standartlarına göre ise, Akademik Benlik Kavramı Ölçeği oldukça yetersiz bulunurken, Gazi Erken Çocukluk Gelişimi Değerlendirme Aracı söz konusu standartların çoğunu kısmen de olsa karşılamıştır. Çüm (2013), benzer bir çalışmayı ölçek geliştirme makaleleri üzerinde yürütmüş ve 2005-2013 yılları arasında TÜBİTAK Ulakbim Ulusal Veri Tabanı'nda psikoloji ve eğitim bilimleri kategorilerinde yer alan dergilerdeki ölçek geliştirme ve uyarlama makalelerinden 50 tanesini tabakalı seçkisiz yöntem ile belirleyerek incelemeye almıştır. Araştırma sonucunda, incelenen 29 ölçek geliştirme çalışmasında inceleme formunda yer alan ölçek geliştirme aşamalarına yaklaşık %67 oranında uygun bilgilerin rapor edildiği, ölçek geliştirme ilkelerine ise %40.19 oranında uygun bilgilerin rapor edildiği belirlenmiştir. Ölçek uyarlama aşamalarına göre incelenen 21 ölçek uyarlama çalışmasında ise, ölçek uyarlama aşamalarına yaklaşık %46 oranında uygun bilgilerin rapor edildiği, ölçek uyarlama ilkelerine ise %27 oranında uygun bilgilerin rapor edildiği belirlenmiştir.

Ülke çapında yapılan ve psikolojik testlerin psikometrik niteliklerine ilişkin çalışmalar incelendiğinde, neredeyse hepsinde benzer sonuçlara ulaşıldığı görülmektedir. Yapılan her çalışmada geliştirilen ya da uyarlanan psikolojik testlerin, eksiklerine, yetersizliklerine vurgu yapılmış ve bu ölçme araçlarının ölçme standartlarından uzak olduğu belirlenmiştir. Özellikle uyarlama çalışmalarının yalnızca hedef dile çeviri olarak kabul gördüğü belirlenmiş ve gerekli izleme çalışmalarının yapılmadığı, testlerin psikometrik niteliklerinin sorgulandığı çalışmaların gerçekleştirilmediği saptanmıştır. Artış gösteren psikolojik testlerin takibini sağlamak ve gereken revizyonları yapmak için bu alanda daha çok bilimsel çalışmaya ihtiyaç duyulduğu ilgili araştırmalarla ortaya çıkmış bir gerçektir. Yüksek lisans ve doktora tezleri kapsamında ölçek geliştirme konuları sıklıkla ele alınsa da, bu çalışmalarda asıl olan yeni bir ölçeğin alana kazandırılması olmaktadır. Bu çalışmalar da ülke çapında geliştirilen ölçme aracı sayısını arttırmakta ve ilgili alana yeni psikolojik testler sunmaktadır, fakat diğer psikolojik testlere ilişkin yaşanan sıkıntılar bu çalışmalar için de geçerli olmaktadır. Geliştirilen bu ölçekler de, diğer psikolojik testler

gibi psikometrik nitelikleri bakımından incelenmeye ve izleme çalışmalarına ihtiyaç duymaktadır. İlgili çalışmalar incelendiğinde ise, hiçbir çalışmanın tez çalışmaları kapsamında geliştirilen ölçme araçlarının psikometrik niteliklerini incelemeye yönelik olmadığı görülmekte ve tüm çalışmaların alanyazında makalelerin ya da hali hazırda kullanılan psikolojik testlerin incelendiği çalışmaların yer aldığı görülmektedir. Ölçek geliştirme amaçlı tez çalışmaları sonucunda da bir ölçek ortaya çıktığı ve geliştirilen bu ölçeğin kullanıma sunulduğu göz önünde bulundurulduğunda, bu ölçeklerin psikometrik niteliklerine ilişkin sorgulamaların ve takip çalışmalarının yapılmasının bir gereklilik olduğu sonucuna ulaşılmaktadır. Özellikle ölçek geliştirme amaçlı tez çalışmalarının sayısı göz önüne alındığında, bu çalışmalar sonucu geliştirilen ölçeklerin hiç de az olmadığı görülmektedir ve bu durum da bu ölçme araçlarına gereken önemin verilmesini zorunlu kılmaktadır. Söz konusu tez çalışmaları kapsamında geliştirilen ölçeklerin nitelikleri hakkında bilgi sahibi olmak için, bu tez çalışmaları incelenmeli ve tez çalışmaları kapsamında geliştirilen ölçeklerin nitelikleri belirlenmeli ve gereken revizyonlar yapılmalıdır.

Bu inceleme hem ölçek geliştirme amaçlı tez çalışmalarının durumu hakkında bilgi sağlamak için hem de geliştirilen ölçeklerin diğer araştırmacılar tarafından tanınmasına, bilinmesine olanak sağlamak için bir ihtiyaçtır. Bu ihtiyaç göz önünde bulundurularak, bu çalışmada eğitim bilimleri alanında gerçekleştirilen ve ölçek geliştirmeyi temel amaç olarak belirten doktora tezlerinden beşi seçkisiz olarak belirlenerek Eğitimde ve Psikolojide Ölçme Standartları'ndan Test ve Ölçek Geliştirme Standartları'na göre incelenecektir. Böylece doktora tez çalışmaları doğrultusunda geliştirilen ölçeklerin mevcut durumları belirlenecek, geliştirilmesi gereken yönleri saptanacaktır. Geliştirilen psikolojik testlerin ölçek geliştirme aşamalarına uygunluğunun incelenmesi hem söz konusu psikolojik testin ölçmeyi amaçladığı psikolojik özelliği daha doğru şekilde ölçmesini sağlayacak, hem de o psikolojik test kullanılarak yapılan çalışmaların daha hedefe yönelik olmasına katkıda bulunacaktır. Dünya çapında geçerliliği kabul edilen standartların Türkiye'de tez çalışmaları kapsamında geliştirilen psikolojik testlerde incelenmesi sağlanarak bu çalışmanın daha sonradan yapılacak çalışmalara örnek oluşturması ve psikolojik testlerin gelişimine katkıda bulunması beklenmektedir. Araştırma kapsamında incelenen ölçek geliştirme çalışmalarının eksik yönlerinin saptanması da bu ölçekleri kullanan araştırmacılara söz konusu ölçekler hakkında öngörü kazandırması umulmaktadır.

Araştırmanın Amacı

Bu çerçevede bu çalışmanın amacı 2009 ve 2014 yılları arasında eğitim bilimleri alanında doktora tezi kapsamında hazırlanan ölçek geliştirme çalışmalarının, APA tarafından geliştirilen test ve ölçek geliştirme standartlarına uygunluğunu sistematik bir biçimde incelemek, belirlenen bu standartları karşılama konusunda eksiklikleri varsa belirleyerek, bu eksikliklerin giderilmesine yönelik önerilerde bulunmaktır.

YÖNTEM

Araştırma Modeli

Bu araştırma, nitel araştırmaların içinde yer alan bir doküman inceleme çalışmasıdır. Nitel araştırmaların doğası gereğince derinlemesine incelemeler ve analizler yapılmakta ve bu nedenle nicel araştırmalarda olduğu gibi büyük örneklemle çalışmak mümkün olmamaktadır. (Yıldırım ve Şimşek, 2011). Bu çalışmada da eğitim bilimleri alanında yapılan doktora tezlerinden tabakalı seçkisiz yöntemle belirlenen beş ölçek geliştirme

konulu doktora tezi test ve ölçek geliştirme standartlarına göre incelenmiş ve söz konusu ölçek geliştirme çalışmaları derinlemesine analiz edilmiştir. Yapılan analizler sonucu tez çalışmaları sonucunda geliştirilen ölçeklerin incelenen standartlara göre durumları ortaya çıkarılmış ve yapılması gereken revizyonlar belirlenmiştir.

Araştırma Grubu

Araştırma kapsamına alınan ölçek geliştirme amaçlı doktora tezlerin belirlenmesi için ülke çapında yapılan yüksek lisans ve doktora tezlerinin yayınlandığı Ulusal Tez Merkezi veri tabanı kullanılmıştır. İncelemeye alınacak tezlerin güncel tez çalışmaları olması gözetilerek tarama 2009-2014 yılları arasında sınırlandırılmıştır. Son beş yıl içinde yapılan ve veri tabanında izinli olarak yayınlanan tüm ölçek geliştirme konulu tezlere ulaşılmıştır. Bir sonraki aşamada bu tezler kendi içinde alanlarına ve derecelerine göre gruplandırılmıştır. Söz konusu veri tabanından temin edilen tezler öncelikle doktora ve yüksek lisans tezi olarak iki kategoriye ayrılmıştır. Derecelerine göre belirlenen doktora tez çalışmaları da eğitim bilimleri ve diğer alanlarda gerçekleştirilen tez çalışmaları olarak gruplandırılmıştır. Eğitim bilimlerinde yapılan doktora tez çalışmaları incelenmiş ve temel amacı ölçek geliştirme olan 11 doktora tez çalışması olduğu saptanmıştır. Belirlenen doktora tez çalışmalarından beşi seçkisiz olarak belirlenerek, araştırma kapsamına alınmıştır. Araştırma kapsamında incelemeye alınan doktora tezlerini hazırlayan araştırmacılar, tezlerin hazırlandığı üniversiteler ve tezlerin isimleri etik ilkeler gözetilerek verilmemiştir. Bu bilgiler yerine söz konusu tez çalışmaları kapsamında geliştirilen ölçeklerin konuları belirtilmiş ve bu ölçekler 1'den 5'e kadar kodlanarak verilmiştir. İncelemeye alınan ölçek geliştirme konulu doktora tezleri Tablo1.'de sunulmaktadır.

Tablo 1. *Ölçek Geliştirme Aşamalarına Göre İncelemeye Alınan Doktora Tezleri*

Tezlerde Yer Alan Ölçeklerin Konusu ve Kodu	Enstitü	Anabilim Dalı	Yılı
1. Çocuklar İçin Geliştirilen Bir Müzik İletişim Ölçeği	Sosyal Bilimler Enstitüsü	Güzel Sanatlar Eğitimi Anabilim Dalı Müzik Eğitimi Bilim Dalı	2010
2. Ergenlerin Uyumunu Değerlendirmek İçin Geliştirilen Bir Ölçek	Eğitim Bilimleri Enstitüsü	Eğitim Bilimleri Ana Bilim Dalı Rehberlik Ve Psikolojik Danışmanlık Bilim Dalı	2013
3. Ergenlerin Çalışma Anlayışına Yönelik Bir Ölçek	Eğitim Bilimleri Enstitüsü	Eğitim Bilimleri Ana Bilim Dalı Rehberlik Ve Psikolojik Danışmanlık Bilim Dalı	2009
4. Bir Derse Yönelik Tutum Ölçeği	Eğitim Bilimleri Enstitüsü	Ortaöğretim Fen ve Matematik Alanları Eğitimi Anabilim Dalı Biyoloji Öğretmenliği Bilim Dalı	2011
5. Çok Engelli Görmeyen Çocukların İçin İletişim Becerilerin Yönelik Bir Ölçek	Eğitim Bilimleri Enstitüsü	İlköğretim Ana Bilim Dalı Okul Öncesi Öğretmenliği Bilim Dalı	2012

Veri Toplama Araçları

Araştırma kapsamında incelenmesine karar verilen ölçek geliştirme konulu beş doktora tezi Ulusal Tez Merkezi veri tabanından indirilerek temin edilmiştir. Araştırma kapsamına alınan tezleri incelemek için kullanılan kontrol listesi, araştırmacı tarafından daha önceki çalışmada (Dirlik, 2013) "Standards for Educational and Psychological Testing (Eğitim ve Psikolojide Ölçme Standartları)" kitabından Türkçeye çevrilmiş ve uzman görüşleri doğrultusunda kısa ve uzun form olarak hazırlanmıştır. Söz konusu standartlar APA tarafından belli zaman aralıklarıyla yayınlanan ve güncellemesi yapılan standartlardır. Bu standartların 1985 yılında yayınlanan versiyonu Türkçeye çevrilmiş ve kitap olarak basılmıştır (Hovardaoğlu ve Sezgin, 1998). Bu çalışma kapsamında daha güncel standartları kullanarak bir kontrol listesi hazırlamak amaçlanmış ve bu nedenle son olarak 1999 yılında yayınlanan APA standartları araştırmacı tarafından Türkçeye çevrilerek kullanılmıştır. Bu çalışma kapsamında hazırlanan kontrol listesinin uzun formu kullanılarak belirlenen doktora tezleri kapsamında geliştirilen ölçekler incelenmiştir. Söz konusu kontrol listesi 28 standardı içermektedir ve standartların ölçek geliştirme tezlerinde karşılama düzeyi 3'lü olarak derecelendirilerek incelenmiştir. İncelenen standart ölçek geliştirme çalışmalarında hiç bulunmuyorsa 1 ile, kısmen karşılanıyorsa 2 ile ve tamamen karşılanıyorsa 3 ile ifade edilmiştir.

Verilerin Analizi

Araştırma kapsamına alınan doktora tezleri hazırlanan kontrol listesi kullanılarak araştırmacı tarafından incelenmiştir. Araştırmanın amacı kapsamında elde edilen verilerin analiz edilmesi için içerik analizi yöntemi kullanılmıştır. İçerik analizi; yazılı veya resimli bir belgede yer alan mesajın nicelleştirilmesi işlemidir (Krippendorf, 2003). Berelson ise içerik analizi yöntemini "İletişimin görünen içeriğinin nesnel, sistematik ve nicel yollardan betimlenmesi" olarak tanımlamıştır (Akt. Tavşancıl ve Arslan, 2001) Bir başka tanım ise, içerik analizini metin, kitap, belge vb. malzemenin belli özelliklerinin belirlenmesi amacıyla yapılan bir tür tarama olarak nitelendirmektedir (Karasar, 2009). Bu çalışmada da, belirlenen psikolojik testler, ölçek geliştirme standartlarına göre analiz edilmiş ve bu çalışmaların standartlara göre durumu oluşturulan kategoriler temel alınarak belirlenmiştir. Dolayısıyla içerik analizi tekniği içinde yer alan kategorisel analiz bu çalışmanın amacıyla örtüşmektedir ve araştırma kapsamına alınan ölçek geliştirme tezleri bu teknik ile incelenmiştir.

Kategorisel içerik analizi; frekans analizi ile birlikte en sık kullanılan içerik analizi türlerinden biridir. Kategorisel içerik analizi, belirli bir iletişim içeriğinin belirli birimlere bölünmesini ve ardından bu birimlerin önceden belirli ölçütlere göre kategoriler halinde gruplandırılmasıdır. Kategoriler, araştırma amacı doğrultusunda farklı şekillerde oluşturulabilir (Tavşancıl ve Arslan, 2001). Bu çalışmadaki kategoriler daha öncede belirtilen ve araştırmacı tarafından hazırlanmış olan kontrol listesinde yer alan test geliştirme standartlarıdır. İçerik analizi yönteminde bu şekilde kurama dayalı olarak geliştirilen kategoriler önsel oluşturulmuş kategoriler (Stemler, 2001) ya da tümden gelişimsel yöntemle oluşturulmuş kategoriler olarak adlandırılır. Bu tür kategori oluşturmada, kategoriler alan uzmanları tarafından belirlenir ve belli bir kurama dayalı olarak geliştirilir (Yıldırım ve Şimşek, 2011). Bu çalışmadaki kategoriler de APA tarafından oluşturulmuş ve ölçek geliştirme aşamaları dikkate alınarak hazırlanmıştır.

İçerik analizi yöntemi ile yapılan incelemelerin güvenilirliğinin belirlenmesi için çeşitli yollara başvurulur ve bu yollardan biri de aynı araştırmacının farklı zamanlarda aynı materyali tekrar kodlaması ve kodlamalar arasındaki tutarlığın hesaplanmasıdır. Bu

çalışmada gerçekleştirilen içerik analizinin güvenilirliğini belirlemek için araştırmacı tarafından söz konusu ölçek geliştirme tezleri üç hafta arayla iki kez incelenmiş ve incelemeler arası tutarlılık, uyuşum yüzdesi formülü kullanılarak hesaplanmıştır (Stemler, 2001; Tavşancıl ve Aslan, 2001).

$$\text{Güvenirlilik} = \frac{\text{uzlaşma sayısı}}{\text{uzlaşma sayısı} + \text{uzlaşmama sayısı}}$$

Formül 1. Puanlayıcılar/ Puanlamalar Arası Tutarlılık Hesaplaması

Yapılan kodlamalar arası uyuşumun belirlenmesinde Formül 1 kullanılmış ve iki kodlama arasındaki uyuşum 0.89 olarak bulunmuştur. Elde edilen sonucun 0.70'ten yüksek olması kabul edilebilir bir uyuşma düzeyi anlamına geldiğinden (Tavşancıl ve Aslan, 2001) incelemeler arası uyuşumun var olduğu, bir diğer ifadeyle bu yöntem için kabul edilebilir düzeyde güvenilirliğin sağlandığı sonucuna ulaşılmıştır.

Elde edilen ham verilerin sayısallaştırılması için frekans dağılımından yararlanılmıştır. Hazırlanan kontrol listesindeki her bir ölçek geliştirme çalışması ve her bir standart için frekans ve yüzde değerleri hesaplanarak tablo halinde sunulmuştur ve böylece her psikolojik testin, kontrol listesindeki standartları karşılama durumu belirlenmiştir. Bu verilerin analizi SPSS 20.0 paket programı kullanılmıştır.

BULGULAR

Araştırma kapsamında incelemeye alınan ölçek geliştirme tezleri, test geliştirme standartları temel alınarak incelenmiştir. Kontrol listesinde yer alan standartlardan sekizi ölçek geliştirme tezlerinde yer almayan durumlara yönelik olduğundan inceleme dışında tutulmuştur. İnceleme dışında tutulan standartlar, açık uçlu maddelerin olması ve maddelerin farklı ağırlıklandırılmış olması durumunda incelemeye alınması gereken standartlardır. Tezler kapsamında geliştirilen ölçeklerin hiçbiri bu tür durumları kapsamadığından bu maddeler tüm tez çalışmaları için inceleme dışında bırakılmıştır. Tez çalışmaları 20 standart gözetilerek incelenmiş, fakat bazı standartlar geliştirilen ölçeklerin özelliğinden dolayı söz konusu çalışmalarda yer almadığı için, her tez çalışması için incelenen standart sayısı farklılaşmıştır.

Bulguların sunulmasında incelemeye alınan tez çalışmaları Tablo 1.'deki sıra numaraları ile verilmiştir ve tüm çalışmaların standartları karşılama durumu özet olarak sunulmuştur. Her tez çalışması için karşılanan, kısmen karşılanan ve karşılanmayan standartların frekansları verilmiş ve bu standartlara ilişkin yorumlar yapılmıştır. Tüm standartların her ölçek geliştirme tezinde karşılanma durumuna ilişkin detaylı tablo ise EK-1 'de verilmiştir.

Doktora Tezlerinin Test ve Ölçek Geliştirme Standartlarına Uygunluğunun İncelenmesi

İncelemeye alınan doktora tezlerin kontrol listesinde yer alan ve bu tezler için incelemeye alınması uygun bulunan test geliştirme ve ölçek standartlarına göre durumu Tablo 2. 'de özetlenmektedir.

Tablo 2. *Test ve Ölçek Geliştirme Standartlarının Belirlenen Beş Doktora Tezinde Yer Alma Durumlarının İncelenmesi*

Test ve Ölçek Geliştirme Standartları					
İncelenen Ölçek Geliştirme Tezleri	Tez 1	Tez 2	Tez 3	Tez 4	Tez 5
Değerler	f	f	f	f	f
1-hiç karşılanmıyor	9	-	6	4	1
2-kısmen karşılanıyor	4	2	5	3	3
3-tamamen karşılanıyor	6	16	5	9	13
Toplam	19	18	16	16	17

Tablo 2 incelendiğinde, daha önce de belirtildiği gibi her ölçek geliştirme tezi için farklı sayıda standardın incelemeye alındığı görülmektedir. Ortak olarak inceleme dışında bırakılan standartlara ek olarak, her ölçek geliştirme tezinde farklı standartların inceleme dışında tutulmuştur.

İncelenen tezlerde standartların karşılanma durumuna bakıldığında ise, Tez 1'de 19 standardın incelendiği ve bu standartların 9'unun tez 1 için hiç karşılanmadığı, 4'ünün kısmen karşılandığı ve 6'sının ise tamamen karşılandığı belirlenmiştir. Tez 2'nin standartlara göre durumu incelendiğindeyse, incelenen 18 standardın 16'sının tamamen karşılandığı, 2'sinin kısmen karşılandığı belirlenmiştir. Bir diğer ölçek geliştirme tezi olan Tez 3 için yapılan incelemede ise, bu ölçek geliştirme çalışması için 16 standardın incelendiği ve bu standartların 6'sı bu ölçek geliştirme çalışması için hiç karşılanmazken, 5'i kısmen, diğer 5'i ise tamamen karşılandığı belirlenmiştir. Tez 4 yapılan incelemede ise, yine 16 standarda göre inceleme yapıldığı ve 9 standardın bu tez için tamamen karşılandığı, 4 standardın hiç karşılanmadığı ve 3 standardın ise kısmen karşılandığı sonucuna ulaşılmıştır. İncelemeye alınan son ölçek geliştirme çalışması içinse, 17 standardın incelendiği, 13 standardın tamamen, 3 standardın ise kısmen karşılandığı ve yalnızca 1 standardın hiç karşılanmadığı sonucuna ulaşılmıştır.

Yapılan incelemeler sonucunda, standartları karşılama durumu en yüksek düzeyde olan ölçek geliştirme çalışmasının Tez 2 olduğu belirlenmiştir. Söz konusu tezde ergenler için bir ölçek geliştirme çalışması gerçekleştirilmiş ve geliştirilen ölçeğin geçerlik ve güvenilirlik çalışmalarına ek olarak, norm çalışmaları da yapılmıştır. Bu ölçek geliştirme çalışmasında kısmen karşılanan standartlar ise testin yönergeleri ile ilişkilidir. Bu konuda araştırmacı standartlarda belirtilen kadar detaylı bilgi vermemiş ve test yönergelerinin uzman görüşüne sunulup sunulmadığını belirtmemiştir. Bu ölçek geliştirme tezinden sonra, standartları karşılama düzeyi yüksek olan ikinci tez ise, 5 nolu tezdır. Bu tez

çalışmasında çok engelli görmeyen çocukların iletişim becerilerini ölçmeye yönelik bir ölçek geliştirilmiştir. Bu ölçek geliştirme çalışması için hiç karşılanmayan standart, ölçek geliştirilme sürecinde yargılarına başvuru uzmanlara ilişkin detaylı bilgi verilmesini gerektiren standarttır. Bu tez çalışması için kısmen karşılanan standartlar ise, madde seçim sürecinin ve maddelerin test dışında bırakılması sürecinde yapılanların belgelendirilmesi ve test yönergelerinin uzman yargısına sunulması gerekliliğiyle ilgili standartlardır. Standartları karşılama düzeyi bakımından en iyi üçüncü olan ölçek geliştirme tezi ise Tez 4'tür. Bu tez çalışmasında bir derse yönelik tutum ölçeği geliştirilmiştir. İncelenen 16 standardın 9'unu tamamen karşılayan bu ölçek geliştirme tezi için karşılanmayan standartlar testin amaçlarının net olarak belirlenmesi, test geliştirme süreçlerinin ayrıntılı olarak rapor edilmesi, yargısına başvuru uzmanların niteliklerinin belirtilmesi ve ilgili alanın kapsamı temsil etme düzeyinin belirlenmesi ile ilgili standartlardır. Ayrıca bu tez için kısmen karşılanan standartlar ise, test puanlarının yorumlanmasında kullanılan ölçütlerin belirlenmesi, madde seçim süreçleri ve deneme uygulaması için kullanılan örneklemin belirtilmesi ile ilgili standartlardır. Söz konusu tez çalışması bu kriterlere ilişkin yüzeysel bilgiler içermekte, detayları vermemekte ve tüm süreci belgelendirmemektedir.

İncelenen tezlerde standartları karşılama düzeyi düşük olan iki tez çalışması yer almaktadır ve bunlardan ilki 1 nolu tez çalışmasıdır. Bu tezde çocuklar için bir müziksel iletişim ölçeği geliştirilmiştir. Bu ölçek geliştirme tezi için incelemeye alınan 19 standardın 9'u hiç karşılanmamıştır. Bu standartlar genel olarak, test geliştirme sürecinde toplanması gereken kanıtlar, test yönergelerinin uzman görüşlerine sunulmuş olması, test puanlarının kullanımında ölçütlerin belirlenmiş olması, testi inceleyen uzmanların nitelikleri ile ilgili belirtilmesi gerekenler, madde seçim sürecinin belirtilmesi ile ilgili kriterler ve puanlayıcıların seçimi ve eğitimi ile ilgili kriterler yer almaktadır. Bu çalışmada kısmen karşılanan standartlar ise, test yönergelerinin açıklığı, test geliştirme sürecinde toplanan kanıtların belirtilmesi ve maddelerin test dışına çıkarılması sürecinde yapılanlar ile ilgili standartlardır. İlgili standartları karşılama düzeyi düşük olan bir diğer tez ise, 3 nolu tezdır ve bu tez kapsamında ergenlerin çalışma anlayışlarını incelemeyi amaçlayan bir ölçek geliştirilmiştir. Bu ölçek geliştirme tezinde incelenen 16 standardın 6'sının bu tez için hiç karşılanmadığı belirlenmiştir. Bu ölçek geliştirme tezi için hiç karşılanmayan standartlar, diğer tezlerde de olduğu gibi, yönergelerin geliştirildiği süreç ve uzman yargılarına sunulması ile ilgili standartlar, test puanlarının yorumlanmasında kullanılan ölçütler ile ilgili standartlar ve görüşleri alınan uzmanların nitelikleri ile ilgili standartlardır. Bu çalışma için kısmen karşılanan standartlar ise, maddelerin geliştirilme, gözden geçirilme ve test dışına çıkarma sürecinde yapılanların rapor edilmesi, madde özelliklerinin belirlenmesinde kullanılan örneklemin seçilme süreci ve yanıtlayıcılar için hazırlanan yönergelerin ayrıntılarının belirtilmesi ile ilgili standartlardır.

Ölçek geliştirme tezlerinin standartlara göre incelenmesi sonucunda, söz konusu tezler için ortak olarak hiç karşılanmayan ya da kısmen karşılanan standartların, genel olarak yönergelerle, madde ve test geliştirme sürecinde elde edilmesi gereken kanıtlarla ve başvuru uzmanlarla ilgili olduğu görülmektedir. İncelenen ölçek geliştirme tezlerinin dördünde, testin amacının net olarak belirtildiği, test edilen psikolojik özelliğin tanımlanmış olduğu, yazılan maddelerin ölçme yapılacak gruba uygun olarak belirlendiği, maddelere ve cevaplama biçimlerine yönelik uzman yargılarına başvurulduğu ve madde seçim sürecinde kullanılan madde özelliklerinin belirtildiği gözlenmiştir. İncelenen beş ölçek geliştirme tezinin dördünde, puanlayıcıların seçimi ve eğitimine ilişkin yapılan

çalışmalardan bahsedilmemiş ve puanlayıcı güvenilirliğini çalışılmamıştır. Ayrıca incelenen 2 nolu ölçek geliştirme tezinde, ölçeğin son formu verilmediği ve yönergeler yer almadığı için ilgili standartlar bu tez için incelenememiştir.

İncelenen Ölçek Geliştirme Tezlerinin Güncellenmesi Gereken Yönleri

Eğitimde ve Psikolojide Ölçme Standartları içinde yer alan ölçek geliştirme standartlarına göre incelenen doktora düzeyindeki ölçek geliştirme tezlerinin, standartları karşılama durumlarına göre güncellenmesi gereken yönleri belirlenmiştir. İncelenen tezlere genel olarak bakıldığında, tüm tez çalışmalarında ölçek geliştirilme sürecinde uzman yargılarına başvurulduğu fakat uzmanlar hakkında verilen bilgilerin standartta belirtildiği kadar detaylı olmadığı belirlenmiştir. Söz konusu standart uzmanların niteliklerinin, alanlarının, o alandaki deneyimlerinin ve demografik özelliklerinin belirtilmesini gerektirmektedir. Ayrıca yalnızca maddeler konusunda değil test yönergeleri konusunda da uzman görüşlerine başvurulmalı ve bu süreç de belgelendirilmelidir. Dolayısıyla incelenen tezler kapsamındaki geliştirilen tüm ölçekler bu standart bakımından revize edilmelidir.

İncelenen tezlerden 1 nolu ve 4 nolu tez için, geliştirilen ölçeğin kapsamı tanımlanmalı ve sınırlandırılmalıdır. Birden fazla uygulayıcı ile uygulanan 1 nolu tez kapsamında geliştirilen ölçme aracına ilişkin puanlayıcı tutarlılığı çalışmaları gerçekleştirilmeli ve puanlayıcıların nasıl eğitildiği belirtilmelidir. İncelenen tezlerden yalnızca ikisinde elde edilen puanların yorumlanmasında kullanılacak ölçütlerin verildiği belirlenmiştir. Dolayısıyla diğer tez çalışmalarında geliştirilen ölçeklerden elde edilen puanların nasıl yorumlanacağı belirtilmeli, puanları yorumlamada kullanılacak ölçütler, normlar net olarak belirlenmelidir.

Çalışma kapsamında incelemeye alınan tezlerde geliştirilen ölçeklere ilişkin gözden geçirilmesi gereken son nokta ise, uygulamalardan kaynaklanabilecek değişiklikler için kullanıcıların bilgilendirilmesidir. İncelenen hiç bir ölçek geliştirme tezinde, geliştirilen ölçekler için uygulamaya ya da testi alanlara göre kabul edilebilir olan değişiklikler belirtilmemiş, testi alanların özel ihtiyaçları (olası engel durumları, acil oluşabilecek fizyolojik ihtiyaçları vb.) göz önünde bulundurulmamıştır. Tüm ölçek geliştirme çalışmalarında bu konu göz önünde bulundurulmalı, özel ihtiyaçları olan katılımcılar için ya da farklı uygulama ortamları için kabul edilebilir değişiklikler ve kabul edilebilir olma nedenleri belirtilmelidir.

SONUÇLAR ve TARTIŞMA

Araştırma kapsamında incelemeye alınan beş doktora tez çalışması kapsamında geliştirilen psikolojik testler, APA tarafından geliştirilen ve araştırmacı tarafından Türkçeye çevrilen Test ve Ölçek Geliştirme Standartları temel alınarak incelenmiştir. İncelenen beş ölçek geliştirme tezinin ikisi ilgili standartları iyi düzeyde karşılarken, biri orta, diğer ikisi ise düşük düzeyde karşılamıştır.

Elde edilen bulgular alan yazınla ilişkilendirildiğinde, 1971 yılında Milli Eğitim Bakanlığı (MEB) Planlama Araştırma ve Koordinasyon Dairesi Rehberlik ve Psikolojik Danışma Bölümü'nce yapılan çalışmada incelenen psikolojik testlerde belirlenen eksiklerin genel olarak bu tez çalışmalarında geliştirilen ölçeklerde de bulunduğu belirlenmiştir. MEB(1971) tarafından gerçekleştirilen bu çalışmada, o tarihlerde kullanılmakta olan 20 psikolojik testin, hiçbirisi için geçerlik, güvenilirlik ve norm çalışmalarının yapılmadığı belirlenmiştir. Psikolojik testlerin psikometrik özelliklerine ilişkin bu eksikliklerden

geçerlik ve norm çalışmalarına ilişkin olanların bu çalışmada incelenen ölçek geliştirme çalışmalarında da geçerli olduğu belirlenmiştir. İncelenen beş tez çalışmasında geliştirilen ölçeklerin yalnızca biri için (Tez 2) normlar geliştirilmiştir. İncelemeye alınan bir diğer tez olan beş nolu tez kapsamında geliştirilen ölçekten elde edilen puanların yorumlanmasına ilişkin bir takım sınır puanlar oluşturulmuş olsa da ölçeğe ilişkin normlar geliştirilmemiştir. Diğer üç tez çalışmasında geliştirilen ölçekler içinse ölçek puanını yorumlamada kullanılacak ölçütler net olarak belirlenmemiştir. Ayrıca incelenen ölçek geliştirme tezlerinin yalnızca ikisi (Tez 2 ve 5) ölçme yapılacak alanın kapsamını tanımlayarak ve uzman görüşüne sunarak, kapsam geçerliğine ilişkin kanıtlar toplamıştır. Bu çalışmadaki sorunlar ile MEB tarafından 1971 yılında gerçekleştirilen çalışmada ortaya çıkan sorunların benzerlik göstermesi, aradan uzun yılların geçmesine rağmen, ölçek geliştirme çalışmalarında aynı sorunların yaşanıyor olduğunu göstermektedir. Başka bir benzer çalışma olan ve araştırmacı tarafından gerçekleştirilen çalışmada ise (Dirlik, 2013), dört psikolojik testi bu çalışmada da kullanılan APA standartlarına göre incelenmiştir. Söz konusu çalışmada incelenen psikolojik testlerin test geliştirme standartlarının çoğunu kısmen karşıladığı belirlenmiştir, fakat puanlayıcılar için yapılması gereken işlemlere, uzmanların nitelikleriyle ilgili belirtilmesi gerekenlere ve madde seçim sürecinde rapor edilmesi gereken durumlara ilişkin standartların psikolojik testlerde ortak olarak karşılanmadığı tespit edilmiştir. Söz konusu durum bu çalışma için de geçerli olmuştur ve incelemeye alınan beş ölçek geliştirme odaklı doktora tezinden yalnızca biri (Tez 2), puanlayıcıların eğitimine ve tutarlılığına ilişkin yapılması gereken çalışmaları belirten standardı tamamen karşılamıştır. Yine incelemeye alınan beş psikolojik testin yalnızca biri (Tez 2), test geliştirme sürecinde görüşlerine başvuru alan uzmanların nitelikleri hakkında detaylı bilgi vermiştir, diğer tezler bu konuda gereken düzeyde bilgi vermemiştir. Bir diğer benzer çalışmada Çüm(2013), incelediği ölçek geliştirme konulu makalelerinin, ölçek geliştirme aşamalarına uygunluğunun %67 olduğunu belirlemiştir. Çüm(2013)'ün çalışmasına benzer şekilde, bu çalışmada incelenen tezler de ölçek geliştirme standartlarını genel olarak orta düzeyde karşılamaktadır.

Özetle, bu çalışmada kapsamında elde edilen bulgular alan yazınla ilişkilidir ve benzer sonuçlar üretmiştir. Fakat bu benzer sonuçların psikolojik testlere ilişkin olumlu bir şekilde yorumlanması mümkün değildir. Farklı zamanlarda yapılan araştırmalarda, farklı psikolojik testlerin benzer eksikleri olduğu belirlenmiştir. Bu durum da ülke çapında geliştirilen özgün psikolojik testlerin sayısında gözlenen artışa rağmen, niteliksel olarak benzer sorunların devam ettiğini göstermektedir. Bu soruna çözüm olarak da farklı araştırmacıların farklı psikolojik testler üzerinde incelemeler yapması, sonuçların da araştırmacılar tarafından takip edilerek psikolojik testler için gereken revizyonların yapılması önerilebilir. Bu çalışmada kullanılan veri toplama aracı olan kontrol listesi kullanılarak hali hazırda kullanılan ölçme araçları incelenebilir ve gereken revizyonlar belirlenebilir. Bu gibi çalışmaların kullanılmakta olan psikolojik testlerin gelişimine katkıda bulunacağı ve test geliştirecek araştırmacılara yol göstereceği düşünülmektedir.

KAYNAKLAR

- American Psychological Association, American Educational Research Association & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington: American National Research Education.
- Anastasi, A. (1988). *Psychological testing*. (6thEd.). New York: MacMillan Publishing Co. Inc.
- Baykul, Y., Turgut, F. (2010). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Pegem Akademi Yayıncılık.
- Bilgin, N. (2006). *Sosyal Bilimlerde İçerik Analizi*. Ankara: Siyasal Kitabevi.
- Cansever, G. (1982). *Klinik Psikolojide Değerlendirme Yöntemleri*. İstanbul: Boğaziçi Üniversitesi Yayınları. Gözlem Matbaacılık Koll. Sti.

- Conway, Malcolm J. (2006). *How to collect data: measurement and evaluation*. NY: American Society for Training and Development.
- Cronbach, L. J. (1990). *Essentials of psychological testing*. (5thEd.). New York: Harper Collins Publishers.
- Çukur, C. S. (1999). Kültürün Psikoloji Bilimindeki Yeri Üzerine Görüşler. *Türk Psikoloji Yazıları*, 1(3), 1-16.
- Çüm. S.(2013). "Türkiye’de Psikoloji Ve Eğitim Bilimleri Dergilerinde Yayımlanan Ölçek Geliştirme Ve Uyarlama Çalışmalarının İncelenmesi." Yayınlanmamış yüksek lisans tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Dirlik, E. (2013). "Eğitim Kurumlarında Kullanılan Psikolojik Testlerin Ölçme Standartlarına Göre İncelenmesi." Yayınlanmamış yüksek lisans tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Doğan, E. M. (2009). "Türkiye’deki Psikolojik Çalışmalarda Kullanılan Testlerin Psikometrik Özelliklerinin İncelenmesi: Kültürel Açından Test Uyarlama Çalışmalar." Yayınlanmamış yüksek lisans tezi, Muğla Üniversitesi Sosyal Bilimler Fakültesi, Muğla.
- Hovardaoglu,S. ve Sezgin, N.(1998) . *Eğitimde ve Psikolojide Ölçme Standartları*. Türk Psikologlar Derneği Yayınları: Ankara
- Karasar, N. (2010). *Bilimsel Araştırma Yöntemleri*.(21. Baskı). Nobel Yayıncılık: Ankara.
- Koç, N. (21-21 Kasım 1986). *Psikolojik Hizmetlerde Psikolojik Ölçme Araçları ve Sorunlar*. Türk Eğitim Derneği X. Eğitim Toplantısı'nda sunulmuş bildiri, Ankara.
- Koç, N. (2008). Eğitimimizde Psikolojik Testlerin(Standart Testlerin) Kullanım Durumu, Sorunlar ve Öneriler. 1. *Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi* . Ankara: Ankara Üniversitesi Eğitim Bilimleri Fakültesi Yayınları. 85-103.
- Krippendorff, K. (2003). *Content analysis: an introduction to its methodology*. (2thEd.). USA: Sage Publications.
- MEB,(1971). *Sekizinci Milli Eğitim Şurası- Esaslar, Raporlar, Kararlar*, Ankara: Milli Eğitim Bakanlığı Yayınları.
- Öner, N. (1994). Güvenirliliği ve/veya Geçerliliği Sınanmış Psikolojik Testler. *Türk Psikoloji Dergisi* , 9(33), 9-18.
- Özge, S. (1981). "Psychological Tests Used In Turkey: A Preliminary Survey." Yayınlanmamış yüksek lisans tezi. İstanbul: Bogaziçi Üniversitesi.
- Özoğlu, S., Ç.(2007). *Eğitimde Rehberlik ve Psikolojik Danışma*. (3. Baskı). Ankara: Ankara Üniversitesi Eğitim Bilimleri Yayınları.
- Ramazan, O. (1988). "Survey For a Source Book on Psychological Tests Used in Turkey." Yayınlanmamış yüksek lisans tezi. İstanbul: Bogaziçi Üniversitesi.
- Tavşancıl, E. ve Aslan, E. (2001). *Sözel, Yazılı ve Diğer Materyaller İçin İçerik Analizi ve Uygulama Örnekleri*. İstanbul: Epsilon Yayın.
- Uçman, P. (1980). Psikologların Görev Tanımı ve Eğitimlerine İlişkin Sorunları Giderebilecek Çözüm Yolları.*Türkiye’de Uygulamalı Psikoloji Ve Sorunları* . Ankara: Psikologlar Derneği Yayını.
- Urbina,S. (2004). *Essentials of psychological testing*. (1st Ed.). NJ: Wiley&Sons, Inc.
- Yıldırım, A., Şimşek, H. (2011). *Sosyal Bilimlerde Nitel Araştırma Yöntemleri*. (8. Baskı). Ankara: Seçkin Yayıncılık.
- <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp> 12. 03.2014 tarihinde alınmıştır.

EXTENDED ABSTRACT

Introduction

In the modern world, psychological tests are used in nearly all of the science fields to get information about peoples' psychological properties. One of the most popular area in which psychological tests are used widely is education. For many educational purposes, education specialists benefit from these tests. Sometimes they need psychological test measuring a different properties and they have to develop a new psychological test in parallel to their research goals. Especially in academic dissertations of researchers, there are so many new developed psychological tests. Of course all of these tests have been developed regarding

the rules of test development, but in order to get valid and reliable decisions by using psychological tests, all of the tests should meet the testing standards. To decide whether or not the psychological tests meet the standards and improve the quality of these tests, new studies investigating the properties of these tests should be conducted. In this study, five doctoral dissertations in the educational field themed of scale development are analyzed based on the scale development standards decided by American Psychological Association (APA). With this analysis, the finding out of the situation of the scales developed in the doctoral dissertations is aimed. By this study, the shortcomings of these scales are pointed out, so the researchers who had developed these tests can get the chance of revising their scales by using these standards. By revising the scales according to the standards, it will be possible to get more reliable and valid results. Also the users of these psychological tests can get more information about the tests.

Methods

This study is a kind of qualitative study and it is document analysis. In this study, five of the doctoral dissertations prepared in educational sciences and published in the last five years are decided randomly and analyzed by using the checklist developed based on the test development standards of APA. Categorical content analysis is used and the frequencies which are met by the scales are defined. Also the scales are analyzed in two different times by the researcher. The agreement between analyses is calculated and it is found more than 0.70 which is the critical agreement value.

Results and Discussion

As a result of this study, it has been found that two of dissertations meet the test and scale development standards so highly, one of dissertations meets these standards moderately and the last two dissertations meet these standards in low level. Also the points which should be revised in these theses according to the standards are defined. This study shows that especially the standards related to specialists to whom asked for their ideas about the items and specifications of scales are not met necessarily by the scales. Furthermore, the raters' reliability is not studied any of the scale development studies. This study shows that the shortcomings of scale development studies which have been defined in previous studies are valid for these studies, too. Except from one study, none of the scale development study has got any norms in order to interpret the points taken from the scale. Also there are same reliability and validity problems with the previous studies and one of them is the raters' reliability. It is not studied in any of the scale development studies analyzed in this study. Just like previous studies, it has been detected that there is lack of knowledge about how the items are selected for the final form of the scales. Shortly, it seems that despite of the increment of the number of scales developed in Turkey, similar problems are continuing and to solve these problems, similar studies in which the properties of the scales are questioned should be done more frequently.

Ekler

EK-1 TEST GELİŞTİRME SÜRECİ STANDARTLARI KONTROL LİSTESİ

STANDARTLAR	İncelenen Ölçek Geliştirme Tezleri				
	Tez 1	Tez 2	Tez 3	Tez 4	Tez 5
1.Test geliştiriciler ve yayıncılar test geliştirme sürecine dayanan yeterli sayıda kanıt toplamış ve rapor etmiş olmalıdır.	2	3	3	1	3
2.Testin amacı / amaçları, test edilen alanın tanımı ve testin özellikleri açıkça belirtilmiştir.	3	3	3	1	3
3.Test yönergeleri geliştirildikleri süreç ve mantıkla beraber rapor edilmiştir.	1	2	1	3	
4.Test puanlarını yorumlamada kullanılan özellikler normatif ya da standardize edilmiş örnekler ya da kullanılan kriterler rapor edilmiştir.	1	3	1	2	3
5. Test yönergeleri alanla ve ölçme ile ilişkili, test dışı uzmanlar tarafından gözden geçirilmiştir.	1	2	1	3	2
6.Test uzmanlar tarafından incelenmişse, bu incelemenin amacı, uzmanların uyguladığı süreç ve bu çalışmanın sonuçları belgelendirilmiştir. Uzmanların nitelikleri, benzer deneyimleri ve demografik özellikleri de belgelendirilmiştir.	1	3	1	1	1
7.Madde türleri, cevaplama biçimleri, puanlama ve test uygulama işlemleri, testin amacına, ölçülecek özelliğe ve testi alması hedeflenen grubunun özelliklerine dayanılarak seçilmiştir.	1	3	3	3	3
8.Testin gözden geçirilme süreci, deneysel analizleri ve maddelere ve cevaplama biçimlerine yönelik uzman yargılarına başvurulmuştur.	3	3	1	3	3
9.Maddelerin geliştirilmesi, gözden geçirilmesi, denenmesi ve madde havuzundan seçilmesinde uygulanan prosedürler belgelendirilmiştir.	3	3	2	2	2
10.Test maddeleri test yönergeleri göre farklı kategorilere ya da alt testlere ayrılmışsa sınıflama prosedürleri ve bu sınıflamanın netliği ile uygunluğu belgelendirilmiştir.	-	3	-	-	3

11. Madde denemeleri ve alan testleri uygulandığında, madde denemeleri için örneklemin seçilme prosedürleri ve örneklemin özellikleri rapor edilmiştir. Bu örneklem testin uygulanacağı örneklemin mümkün olabildiğince temsilcisidir.	3	3	1	2	3
12. Test geliştiriciler, testin kapsam alanının tanımlanan alanı ve testin özelliklerini ne oranda temsil ettiğini belirtmişlerdir.	1	3	2	1	3
13. Madde seçimi içerik ya da teorik temeller yerine maddelerin deneysel ilişkilerine dayanıyorsa, test geliştiriciler çapraz geçerlik çalışmaları uygulamış olmalıdır. Farklı çalışmaların hangi oranda aynı madde setini belirttiği belgelendirilmiştir.	1	3	-	-	-
14. Madde özelliklerini kestirmek için kullanılan örneklem tanımlanmıştır.	3	3	2	3	3
15. Madde seçiminin hangi süreçle yapıldığı, madde seçiminde hangi verilerin kullanıldığı, madde güçlüğü, madde ayırt ediciliği gibi, rapor edilmiştir.	3	3	3	3	3
16. Test geliştirmede madde parametreleri için madde tepki kuramı kullanıldıysa, tahmin prosedürleri, model uyumunun kanıtları belgelendirilmiştir.	BU ÇALIŞMADAKİ ÖLÇEK GELİŞTİRME TEZLERİ İÇİN İNCELEME DIŞINDA TUTULMUŞTUR.				
17. Test puanları farklı şekilde ağırlıklandırılmış maddelerden türetiliyorsa, test geliştiriciler bu ağırlıklandırmanın mantığı ve madde ağırlıklarını geliştirme, gözden geçirme ve maddeleri test kapsamından çıkarma sırasındaki prosedürleri rapor etmiştir.	BU ÇALIŞMADAKİ ÖLÇEK GELİŞTİRME TEZLERİ İÇİN İNCELEME DIŞINDA TUTULMUŞTUR.				
18. Madde ağırlıkları deneysel verilerden elde edilirse, bu işlemde kullanılan örnekleme oldukça geniş ve hedef kitlenin temsilcisi olmalıdır.	BU ÇALIŞMADAKİ ÖLÇEK GELİŞTİRME TEZLERİ İÇİN İNCELEME DIŞINDA TUTULMUŞTUR.				
19. Madde ağırlıkları uzman yargılarına dayanıyorsa, yargılamayı yapanların nitelikleri belgelendirilmiştir.	BU ÇALIŞMADAKİ ÖLÇEK GELİŞTİRME TEZLERİ İÇİN İNCELEME DIŞINDA TUTULMUŞTUR.				
20. Testi cevaplayanların açık cevaplı maddelerdeki performansını puanlarken kullanılacak kriterler belirtilmiştir.	BU ÇALIŞMADAKİ ÖLÇEK GELİŞTİRME TEZLERİ İÇİN İNCELEME DIŞINDA TUTULMUŞTUR.				
21. Eğer madde sayısının azaltılması ya da alttestlerin ayrı formlara ayrılması yoluyla testin kısa versiyonu hazırlanmışsa, bu kısa versiyonun yönergeleri de orijinal testle mümkün olduğunca benzer şekilde hazırlanmıştır.	BU ÇALIŞMADAKİ ÖLÇEK GELİŞTİRME TEZLERİ İÇİN İNCELEME DIŞINDA TUTULMUŞTUR.				
22. Maddeleri test kapsamı dışına çıkarma sürecinde kullanılan prosedürler belgelendirilmiştir.	2	3	2	3	2

23. Zaman sınırlaması olan testlerde, test geliştirme araştırmaları, puanlamanın ne dereceye kadar hız bileşenini içerdiği ve bunun ne ölçüde uygun olduğu incelenmiştir.

BU ÇALIŞMADAKİ ÖLÇEK GELİŞTİRME TEZLERİ İÇİN İNCELEME DIŞINDA TUTULMUŞTUR.

24. Testin uygulanması için yönergeler yeterince açık şekilde ve vurgulanarak sunulmuştur.

2 Belirtilmemiştir. 3 3 3

25. Testi cevaplayanların, test geliştirenin amaçları doğrultusunda cevap verebilmesi için cevaplayıcılar için hazırlanan yönergeler yeterince ayrıntılıdır. Uygun durumlarda materyal ve alıştırma örnekleri ya da soru örnekleri verilmiştir.

2 Belirtilmemiştir. 2 3 3

26. Test geliştirici testin uygulama koşullarının bir test cevaplayıcıdan diğerine ya da bir gruptan diğerine değişebileceğini belirtiyorsa, kabul edilebilir değişiklikler belirtilmiş ve bu farklı koşulların kabul edilebilir olma sebepleri belgelendirilmiştir.

BU ÇALIŞMADAKİ ÖLÇEK GELİŞTİRME TEZLERİ İÇİN İNCELEME DIŞINDA TUTULMUŞTUR.

27. Puanlayıcıların seçimi, eğitimi, ve puanları niteleme süreçleri test. geliştirici tarafından rapor edilmiştir.

1 3 Puanlayıcı yok. - -

28. Puanlayıcı güvenilirliği ve puanlayıcıların süre sonunda puanlama standartlarında potansiyel kayışlarının ölçülmesi test geliştirici tarafından rapor edilmiştir.

1 3 Puanlayıcı yok. - -