

PAPER DETAILS

TITLE: STANDART TESTLERE YÖNELİK İNANÇ ÖLÇEĞİ`NİN (STYİÖ) TÜRKÇE UYARLAMASI:
GEÇERLİK VE GÜVENİRLİK ÇALIŞMASI

AUTHORS: Mustafa İlhan,Bayram Çetin,İsmail Kinay

PAGES: 161-189

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/183358>

STANDART TESTLERE YÖNELİK İNANÇ ÖLÇEĞİ'NİN (STYİÖ) TÜRKÇE UYARLAMASI: GEÇERLİK VE GÜVENİRLİK ÇALIŞMASI¹

Mustafa İLHAN

Dicle Üniversitesi, Ziya Gökalp Eğitim Fakültesi
mustafailhan21@gmail.com

Bayram ÇETİN

Gazi Üniversitesi, Gazi Eğitim Fakültesi
bcetin27@gmail.com

İsmail KINAY

Dicle Üniversitesi, Ziya Gökalp Eğitim Fakültesi
ismailkinay84@gmail.com

Özet

Bu araştırmanın amacı, Magee ve Jones (2012) tarafından geliştirilen Standart Testlere Yönelik İnanç Ölçeğini (STYİÖ) Türkçeye uyarlamak ve ölçeğin psikometrik özelliklerini incelemektir. Araştırma 2013-2014 öğretim yılı güz döneminde 361 öğretmen ve 192 öğretmen adayı olmak üzere toplam 553 katılımcıdan oluşan üç ayrı çalışma grubu üzerinde yürütülmüştür. STYİÖ'nün Türkçeye uyarlanması sürecinde öncelikle ölçeğin dilsel eşdeğerliği incelenmiştir. Dilsel eşdeğerlik çalışmasında ölçeğin Türkçe ile İngilizce formundan elde edilen puanlar arasında pozitif yönde, güçlü ve anlamlı korelasyonlar saptanmıştır. Doğrulayıcı faktör analizi (DFA) sonucunda, STYİÖ'nün Türkçe formunun orijinal forma paralel üç faktörlü bir yapıya sahip olduğu ortaya çıkmıştır. STYİÖ'nün güvenilirliği; DFA'dan elde edilen faktör yükleri ve hata varyansı değerlerine dayalı olarak hesaplanan bir güvenilirlik katsayısı olan bileşik güvenilirlik yöntemi ile hesaplanmıştır. Güvenirlik analizi sonucunda, öğretmenlerden toplanan veriler üzerinden hesaplanan güvenilirlik katsayısı .66 ve öğretmen adaylarından toplanan veriler üzerinden elde edilen güvenilirlik katsayısı .77 olarak bulunmuştur. Madde analizinden elde edilen bulgular, ölçekte yer alan maddelerin tamamının hem öğretmenler hem öğretmen adayları için ayırt edici sonuçlar verdiğini göstermiştir. Araştırmadan elde edilen bulgulara dayanarak, STYİÖ'nün Türkçe formunun öğretmenlerin ve öğretmen adaylarının standart testler hakkındaki inançlarını belirlemede kullanılabilecek geçerli ve güvenilir bir ölçme aracı olduğu söylenebilir.

Anahtar Kelimeler: Standart testler, Standart testlere yönelik inanç ölçeği, Geçerlik, Güvenirlik

THE TURKISH ADAPTATION OF BELIEFS ABOUT STANDARDIZED TESTS SCALE (BASTS): THE VALIDITY AND RELIABILITY STUDY

Abstract

The purpose of the present study is to adapt Beliefs About Standardized Tests Scale (BASTS), which is developed by Magee and Jones (2012) to Turkish, and to assess its

¹ IV. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi'nde sözlü bildiri olarak sunulmuştur.

psychometrics properties. Research was conducted on three different study groups, who consisted of 361 teachers and 192 prospective teachers during 2013-2014 education year first semester. In Turkish adaptation process of BASTS, its language equivalence was assessed and strong, positive and significant correlations were obtained between Turkish and English versions. Confirmatory Factor Analysis (CFA) showed three factors for Turkish version of BASTS scale similar to its original version. Reliability of BASTS was calculated with composite reliability which is a method based on factor loadings and errors variance obtained from CFA. As a result of reliability analysis, it was found that the composite reliability coefficients were .66 and .77 for teachers and prospective teachers respectively. The findings of the item analyses showed that all of the items in the scale were discriminatory for both teachers and prospective teachers. Based on these results, it can be concluded that Turkish version of BASTS is a valid and reliable instrument in assessing teachers' and prospective teachers' beliefs about standardized tests.

Key Words: Standardized tests, Beliefs about standardized tests scale, Validity, Reliability

Giriş

İnançlar; birey tarafından doğru olarak kabul edilen bilişsel şemalar olup yaşamda karşılaştığı olay, olgu, kişi ya da nesneleri anlamlandırmasında bireye rehberlik etmektedir (Krows, 1999; Pajares, 1992). İnançların bireyin amaç, düşünce ve davranışlarının temel belirleyicisi olması; öğrenmeye yönelik inanç, epistemolojik inanç ve öz yeterlik inancı gibi birçok farklı inancın öğrenme ve öğretme süreçleri açısından dikkate alınmasını zorunlu kılmaktadır (Ajzen, 2005; Deryakulu, 2006). Öğrenme-öğretme sürecinde dikkate alınması gereken inançlardan biri de değerlendirmeye yönelik inançlardır (İlhan, Çetin & Bars, 2013).

Değerlendirmeye yönelik inançlar; değerlendirmenin öğrenme-öğretme sürecindeki yeri ile ilgili inançlar ve öğrenme-öğretme sürecinde kullanılan değerlendirme yöntemleri ile ilgili inançlar gibi değişik açılardan ele alınabilmektedir (İlhan, Çetin & Bars, 2013). Değerlendirmenin öğrenme-öğretme sürecindeki yeri ile ilgili inançlar; değerlendirme yapmanın gerekliliği, değerlendirmenin öğrenmeyi geliştirip geliştirmediği şeklindeki inançlardan oluşmaktadır (Brown & Hirschfeld, 2008). Öğrenme-öğretme sürecinde kullanılan değerlendirme yöntemleri hakkındaki inançlar için ise; katılımcı değerlendirmeye yönelik inançlar (Brew & Riley, 2011) ve standart testlere yönelik inançlar (Magge & Jones, 2012) olmak üzere iki başlıklı bir sınıflandırma kullanılabilmektedir.

Katılımcı değerlendirmeye yönelik inançlar; öğrencilerin bireysel olarak ya da gruplar halinde sınıf arkadaşlarını değerlendirdikleri (Falchikov, 1995) ve birbirlerine çalışmalarının niteliği hakkında dönütler verdikleri akran değerlendirme (Paul, Owence, Chrispen & Alexander, 2013; Van den Berg, Admiraal, & Pilot, 2006), sınıf içerisinde işlenen konular hakkında öğrencilerin görüşlerini rahat bir biçimde dile getirmesine imkân tanıyan sınıf içi tartışmalar, öğrencilerin kendi değerlendirmesinde aktif rol aldığı (Boud, 1992; Pope, 2001), kendi başarıları ile ilgili karar alma süreçlerine dâhil olduğu öz değerlendirme (Boud & Falchikov, 1989)

*Standart Testlere Yönelik İnanç Ölçeği'nin (STYİÖ)
Türkçe Uyarlaması: Geçerlik ve Güvenirlik Çalışması*

ve öğrencilerin işbirlikli olarak çalışmaları sonucunda ortaya çıkan ürünlerin değerlendirilmesi esasına dayanan grup değerlendirme (Gültekin, 2012) hakkındaki inançlardan oluşmaktadır.

Değerlendirmeye yönelik inançların, öğrenme-öğretme sürecinde kullanılan değerlendirme yöntemleri açısından ele alınması sonucu ortaya çıkan bir diğer başlık standart testlere yönelik inançlardır. Standart testler; yoğun deneysel çalışmaların ürünü olarak geliştirilen, belirli amaçlar doğrultusunda uzman kişiler tarafından hazırlanan (Koç, 1985), uygulama esasları, nasıl puanlanacağı ve puanların yorumlanmasında nasıl bir yol izleneceği sabit olan ölçme araçlarıdır (Cronbach, 1970; Shipley & McAfee, 2009). Diğer bir deyişle standart testler, ne zaman ve nerede değerlendirildiklerinden bağımsız olarak bütün öğrencilere aynı şekilde uygulanmakta, puanlanmakta ve yorumlanmaktadır (Koretz, 2008; Silver & Kenney, 1995). Standart olmayan testler, tek bir sınıftaki ya da tek bir gruptaki öğrencilere uygulanmak üzere hazırlanırken; standart testler bir bölgedeki ya da ülkedeki bütün okullar gibi oldukça geniş gruplara uygulanmak üzere hazırlanmaktadır (Airasian, 2005).

Norm referanslı testler olarak da adlandırılan (Putnam, 2005), standart testlerin uygulanmasının temel amacı; değişik sınıf ve okullardan hatta farklı ülkelerden öğrencilerin birbirleriyle karşılaştırılmasına imkân tanıyacak açık ve adil göstergeler elde etmektir (Özgüven, 2011). Standart testlerin farklı öğrenciler arasında karşılaştırma yapmayı mümkün kılması bu testlerin en önemli avantajıdır (Olswang & Bain, 1988). Standart testlerden, ulusal ve yerel düzeylerde elde edilen normlara dayalı olarak herhangi bir öğrencinin test puanı, tanımlanmış grupların puanları ile karşılaştırılabilmektedir (Gronlund & Linn, 1990; Karet & Hubbell, 2003). Örneğin, ülke çapında uygulanan standart bir testten elde edilen sonuçlara dayalı olarak, ülkedeki farklı şehirler başarı düzeyleri açısından karşılaştırılabilmektedir. Standart testlerin diğer avantajları aşağıdaki gibi sıralanabilir.

- Standart testlerin uygulanması ve puanlanmasına ilişkin yönergeler açık bir şekilde tanımlanmış olduğundan, bu testler farklı uygulayıcılar tarafından aynı şekilde uygulanabilmektedir.
- Standart testler, uzman kişiler tarafından geliştirildiğinden, bu testlerde yer alan maddelerin nitelikleri yüksektir (Gronlund & Linn, 1990).
- Standart testlerde güvenilirlik yüksektir ve bu testler yeterli düzeyde geçerli sonuçlar vermektedir (King, 2008).
- Standart testler genel olarak çoktan seçmeli soru türünde hazırlandığından (Benjamin, 2012) ve bu testlerin uygulanması ile puanlanmasına ilişkin yönergeler açık bir biçimde tanımlanmış olduğundan standart testler kolay bir biçimde uygulanabilmektedir (Stein-Rubin & Fabus, 2012).

Standart testlerin avantajlarının yanında birtakım dezavantajları da bulunmaktadır. Standart testlerin eleştirilmesinin temel nedenlerinden biri; bu testlerin bütün öğrencilerin aynı öğretimi aldıklarını kabul etmesi ve öğrencilerin kişisel, kültürel ve dilsel özelliklerindeki farklılıkları dikkate almamasıdır. Öğrencilerin kültürel ve dilsel özelliklerindeki farklılıkların göz ardı edilmesi standart testlerde yanlılığa neden olabilmektedir (Aksu Ataç, 2012; Gullo, 2005). Standart testlerin diğer dezavantajları aşağıdaki şekilde ifade edilmektedir.

- Standart testler, öğrencinin öğrenme sürecindeki gelişimine yeterince önem vermemekte ve daha çok öğrencilerin birbirleri ile karşılaşılmasına odaklanmaktadır (Marston, Deno & Tindal, 1983).
- Öğrencilerin standart testlerden aldıkları puanlar, yalnızca diğer öğrencilerin performansları ile ilişkili olarak yorumlanmaktadır (Marston, Deno & Tindal, 1983; Meisels, Xue & Bickel, 2001).
- Standart testlerin hazırlanması ve uygulanması yüksek maliyet gerektiren bir süreçtir (Nichols & Nichols, 2005; Walvoord, 2010).
- Standart testlerden elde edilen sonuçlar öğrencinin öğrenme eksikliklerinin neler olduğuna ve nasıl giderilebileceğine yönelik yeterli bilgi içermemektedir (Walvoord, 2010). Dolayısıyla standart testlerden elde edilen sonuçlara dayalı olarak öğrencilere anlamlı geri dönütler verilememektedir (Nichols & Nichols, 2005). Yine aynı nedenle, standart testler düzey belirlemeye yönelik değerlendirmeler için uygun bir ölçme aracı iken; biçimlendirmeye ve yetiştirmeye yönelik değerlendirmelerde standart testlerin uygun bir tercih olmadığı düşünülmektedir.
- Standart testler genellikle çoktan seçmeli sorulardan oluştuğundan, bu tür testler öğrencilerin yaratıcı zekâlarının geliştirilmesinde yeterli olamamaktadır (Başol Göçmen, 2004).
- Standart testlerde yalnızca doğru cevaba odaklanılmaktadır (Aksu Ataç, 2012). Dolayısıyla bu testler öğrencinin öğrenmeye ilişkin çabalarından çok öğrenme süreci sonunda ortaya çıkan ürün ile ilgilenmektedir.
- Standart testler, öğrencilerin performanslarındaki gerçek değişimleri yansıtmada (Marston, Deno & Tindal, 1983) ve üst düzey düşünme becerilerini ölçmede yetersiz kalmaktadır (Dikli, 2003).

Standart testlerin söz konusu avantajları ve dezavantajlarından hareketle, standart testlere yönelik inançlar; standart testler kullanılarak gerçekleştirilen değerlendirmelerin geçerliğine, öğrencilerin ne kadar iyi öğrendiklerini ve yeteneklerini belirlemede standart testlerin yeterli olup olmadığına, standart testlerden alınan puanların yanlılığına ve bu testlerden alınan puanlara dayalı olarak yargıda bulunmanın doğruluğuna ilişkin içsel kabuller şeklinde tanımlanmaktadır (Magge & Jones, 2012).

*Standart Testlere Yönelik İnanç Ölçeği'nin (STYİÖ)
Türkçe Uyarlaması: Geçerlik ve Güvenirlik Çalışması*

Araştırmanın Amacı

Öğretmenlerin inançları; sınıf ortamında aldıkları kararları, öğrenme-öğretme sürecindeki uygulamalarını ve eğitimsel reformlara uyum sağlama düzeylerini etkilediğinden; öğretmenlerin eğitime ilişkin inançlarının belirlenmesi oldukça önemli bir konu haline gelmektedir (Higgins & Moseley, 2001; Luft & Roehrig, 2007; Mansour, 2009). Epistemolojik inanç (Deryakulu & Büyüköztürk, 2005; Schommer, 1990), zekaya yönelik inanç (Abd-El-Fattah & Yates, 2006; İlhan & Çetin, 2013), öğrenmeye yönelik inanç (Bay vd., 2013, Woolley, Benjamin & Woolley, 2004) ve katılımcı değerlendirmeye yönelik inanç (Brew & Riley, 2011; İlhan, Çetin & Bars, 2013) gibi eğitime ilişkin farklı inançları belirlemeye yönelik olarak hem uluslar arası hem de Türkçe alanyazında çeşitli ölçme araçlarının bulunmaktadır. Standart testlere yönelik inançların belirlenmesi söz konusu olduğunda; uluslararası alanyazında bu amaçla kullanılabilecek bir ölçeğin (Magge & Jones, 2012) bulunduğu; Türkçe alanyazında ise böyle bir ölçme aracına rastlanmadığı belirlenmiştir. Standart testlere yönelik inançlar, öğretmenlerin eğitime ilişkin inançlarının önemli bir bileşeni olduğundan (Sikka, Nath & Cohen, 2007), standart testlere yönelik inançları ölçmede kullanılabilecek bir ölçeğin Türk kültürüne kazandırılması önem arz etmektedir. Bu kapsamda araştırmada, Magge ve Jones (2012) tarafından geliştirilen Standart Testlere Yönelik İnanç Ölçeğinin (STYİÖ) Türkçeye uyarlanması ve ölçeğin psikometrik özelliklerinin incelenmesi amaçlanmaktadır.

Yöntem

Bu araştırma, bir ölçek uyarlama çalışmasıdır. Bu bölümde; çalışma grubu, araştırmada kullanılan veri toplama araçları ve ölçeğin Türkçeye uyarlanması süreci hakkındaki bilgilere yer verilmiştir.

Çalışma Grubu

Bu araştırma, 2013-2014 öğretim yılı güz döneminde toplam 553 katılımcıdan oluşan üç ayrı çalışma grubu üzerinde yürütülmüştür. Birinci çalışma grubu, Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi İngilizce Öğretmenliği programına kayıtlı 40'ı (%67.80) bayan ve 19'u (%32.20) erkek olmak üzere toplam 59 öğretmen adayından oluşmaktadır. Bu gruptaki öğretmen adaylarının 19'u (%32.20) 2. sınıfa ve 40'ı (%67.80) 4. sınıfa devam etmektedir. Grubu oluşturan öğretmen adaylarının yaşları 18 ile 32 arasında değişmekte olup grubun yaş ortalaması 22.80'dir. Ölçeğin dilsel eş değerlik çalışması bu gruptan elde edilen veriler üzerinden yürütülmüştür. İkinci çalışma grubu Diyarbakır il merkezindeki okullarda görev yapan 160'ı (%44.30) bayan ve 201'i (%55.70) erkek olmak üzere toplam 361 öğretmenden oluşmaktadır. Grupta 24 farklı branştan öğretmen bulunmakla birlikte, katılımcıların yaklaşık 1/3'i sınıf öğretmenidir. Öğretmenlerin yaşları 22 ile 60 arasında değişmekte olup grubun yaş ortalaması 32.33'tür. Üçüncü çalışma grubu ise, Diyarbakır ilinde bir dershanenin Kamu Personeli Seçme Sınavı (KPSS) kursuna kayıtlı 87'si bayan (%65.40) ve 46'sı (%34.60) erkek olmak üzere

toplam 133 öğretmen adayından oluşmaktadır. Bu grupta 18 farklı branştan öğretmen adayı bulunmaktadır. Katılımcıların yaşları 19 ile 39 arasında değişmekte olup grubun yaş ortalaması 23.81'dir. İkinci ve üçüncü çalışma grubundan elde edilen veriler üzerinden ölçeğin yapı geçerliği, bileşik güvenirliği ve madde analizleri incelenmiştir.

Veri Toplama Araçları

STYİÖ, Magee ve Jones (2012) tarafından üniversite öğrencilerinin standart testler hakkındaki inançlarını belirlemek amacıyla geliştirilmiş kendini rapor etme türünden bir ölçme aracıdır. 5'li likert tipi bir derecelendirmeye sahip olan ölçekte 9 madde bulunmaktadır. Ölçekteki 2, 4 ve 5 numaralı maddeler tersten puanlanmaktadır. Ölçeğin yaklaşık uygulama süresi beş dakikadır. Ölçek; standart testlerin objektifliği hakkındaki inançlar, standart testlere dayalı olarak alınan kararlar hakkındaki inançlar ve standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar olmak üzere üç boyutlu bir yapıya sahiptir. Standart testlerin objektifliği hakkındaki inançlar boyutuyla, bireylerin standart testlerin yanlılığına ve gerçek bilgiyi ölçmek için standart testlerin uygun bir araç olup olmayacağına yönelik inançlarının ölçülmesi hedeflenmektedir. Bu boyutta 5 madde yer almaktadır. Bu boyuttaki maddelere örnek olarak, *“Standart bir testin gerçekten yansız olması imkânsızdır.”* gösterilebilir. Standart testlere dayalı olarak alınan kararlar hakkındaki inançlar boyutu, öğrenciler hakkında alınacak önemli kararlarda standart testlerin uygun bir tercih olup olmayacağına yönelik inançları ölçmeye yöneliktir. Bu boyutta 2 madde bulunmaktadır. Bu boyuttaki maddelere örnek olarak, *“Üçüncü sınıfın üstündeki tüm sınıflarda, kendi sınıf düzeyindeki standart testlerde başarılı olmayan öğrenciler sınıf tekrarı yapmalıdır.”* maddesi gösterilebilir. Üçüncü ve son alt ölçek olan, standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar boyutuyla, standart testlerden elde edilen sonuçlara dayalı olarak ödüllendirme yapmanın doğru olup olmadığına ilişkin inançların ölçülmesi amaçlanmaktadır. Bu boyutta 2 madde bulunmaktadır. Bu maddelere örnek olarak, *“Standart testlerden en yüksek puan alan öğrencilerin öğretmenleri para ödülü almalıdır.”* verilebilir.

STYİÖ'nün orijinal formunun faktör yapısı ve güvenirliği 142 üniversite öğrencisinden oluşan bir araştırma grubu üzerinde incelenmiştir. STYİÖ'nün yapı geçerliği faktör analizi ile belirlenmiştir. Faktör analizi sonucunda, ölçekteki maddelerin toplam varyansın %85'ini açıklayan üç boyut altında toplandığı ve tüm maddelerin faktör yüklerinin .30'un üzerinde olduğu saptanmıştır. STYİÖ'nün güvenirliğinin belirlenmesinde Cronbach Alpha iç tutarlılık katsayısı kullanılmıştır. Cronbach Alpha güvenirlik katsayıları; ölçeğin geneli için .83 ve standart testlerin objektifliği hakkındaki inançlar boyutu için .76 olarak hesaplanmıştır. Standart testlere dayalı olarak alınan kararlar ve standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar boyutlarında yalnızca 2'şer madde yer aldığından, bu boyutların güvenirliğini ortaya koymak için iç tutarlılık katsayısı yerine maddeler arasındaki korelasyon katsayılarından yararlanılmıştır. Standart testlere

*Standart Testlere Yönelik İnanç Ölçeği'nin (STYİÖ)
Türkçe Uyarlaması: Geçerlik ve Güvenirlik Çalışması*

dayalı olarak alınan kararlar hakkındaki inançlar boyutu için iki madde arasındaki korelasyon $r=.65$, $p<.001$ ve standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar boyutu için iki madde arasındaki korelasyon $r=.71$, $p<.001$ olarak bulunmuştur.

İşlem

STYİÖ'nün Türkçeye uyarlanması sürecinde 2 Eylül 2013 tarihinde Robert G. Magee ile iletişime geçilmiş ve ölçeğin uyarlanabileceğine ilişkin gerekli izin alınmıştır. Gerekli izin alındıktan sonra ölçeğin kaynak dil olan İngilizceden hedef dil olan Türkçeye çeviri sürecine geçilmiştir. Çeviri işleminde; Brislin, Lonner ve Thorndike (1973) tarafından önerilen basamaklar takip edilmiştir. Buna göre, STYİÖ'nün Türkçeye çeviri süreci beş adımda tamamlanmıştır.

1. Adım: Kaynak Dilden Hedef Dile Çeviri

Ölçeğin çeviri sürecinin birinci adımını oluşturan kaynak dilden hedef dile çeviri aşamasında, üç İngiliz dil bilimci görev almıştır. Bu aşamada görev alan çevirmenlerin standart testler ile ilgili herhangi bir bilgisinin olmamasına özen gösterilmiştir. Bu şekildeki bir uygulama ile çevirmenlerin yalnızca dilsel kavramlara ve özelliklere odaklanması amaçlanmıştır. Çeviri işleminden sonra, yapılan çeviriler karşılaştırılarak her bir madde için o maddeyi en iyi temsil ettiği düşünülen Türkçe ifadeler belirlenmiştir.

2. Adım: Uzmanlarından Görüş Alınması

Bu basamakta, bir önceki adımda elde edilen Türkçe form, iki ölçme değerlendirme ve standart testler hakkında bilgi sahibi olan iki eğitim programları ve öğretim uzmanı olmak üzere dört uzman tarafından incelenmiştir. Uzmanlar, ölçme ve değerlendirme kavramlarına ilişkin Türkçe terminoloji ile standart testlerin tanımını ve özelliklerini göz önünde bulundurarak, her bir maddeyi anlaşılabilirlik ve Türk kültürüne uygunluk açısından değerlendirmiştir.

3. Adım: Geri Çeviri

İkinci adımda yapılan değerlendirmeler doğrultusunda gözden geçirilen Türkçe form, birinci basamakta görev alan çevirmenlerden farklı iki İngiliz dil bilimci tarafından tekrar İngilizceye çevrilmiştir. Bu adım, maddelere ölçeğin orijinal formundakinden farklı bir anlam yüklenmediğinin ve çeviri sürecinde ölçeğin değişikliğe uğramadığının ortaya konulabilmesi açısından oldukça önemlidir.

4. Adım: Ölçek için Alan Uzmanlarının Görüşlerinin Alınması ve Görüşler Doğrultusunda Gerekli Düzenlemelerin Yapılıp Ölçeğe Son Şeklinin Verilmesi

Geri çeviri sonucunda elde edilen İngilizce ölçek, ölçeğin özgün formu ile karşılaştırıldıktan ve iki form arasında önemli bir farkın bulunmadığı belirlendikten sonra, dördüncü adıma geçilmiştir. Bu adımda, 3'lü derecelendirilmeye sahip bir ölçek kullanılarak, iyi düzeyde İngilizce bilen iki ölçme değerlendirme uzmanının çevirinin uygunluğuna ilişkin görüşleri alınmış ve uzmanlar arasındaki uyum basit

uyum yüzdesi ile kappa istatistiği cinsinden hesaplanmıştır. Uzmanlardan biri ölçekteki 9 maddeden 7'sinin tamamıyla eş değer, 2'sinin ise kısmen eş değer olduğunu belirtmiştir. Diğer uzman, 9 maddeden 8'inin tamamıyla eş değer 1'nin ise kısmen eş değer olduğunu rapor etmiştir. İki uzmanın görüşleri arasındaki uyumu ortaya koymak için hesaplanan basit uyum yüzdesi ve kappa istatistiği değerleri sırasıyla %88 ve .63 olarak bulunmuştur. Basit uyum yüzdesi için %75 değeri alt sınır olarak kabul edilmektedir (Graham, Milanowski & Miller, 2012). Kappa istatistiği için ise .40'ın altındaki değerler kötü uyumu, .40 ile .59 arasındaki değerler zayıf uyumu, .60 ile .74 arasındaki değerler iyi uyumu, .75 ve üzerindeki değerler ise mükemmel uyumu yansıtmaktadır (Landis & Knoch, 1977). Bu ölçütlere göre, uzmanlar arasındaki uyumun iyi olduğu söylenebilir.

Uzmanlar arasındaki uyumun yeterli olduğu belirlendikten sonra, kısmen eş değer olduğu belirtilen maddelerin ifade ediliş şekillerinde uzmanlardan alınan görüşler doğrultusunda bazı değişiklikler yapılmıştır. Örneğin, uzmanlardan biri *“Öğrenciler çok farklı özelliklere sahip oldukları için tek bir standart test onların yeteneklerini ölçmede gerçekten yararlı olmaz”* şeklinde ifade edilen maddenin *“Öğrenciler çok farklı özelliklere sahip olduğundan öğrencilerin yeteneklerini ölçmede standart testler yararlı olmaz”* biçimde değiştirilmesini önermiş, bu değişiklikle maddedeki vurgunun *gerçekten* kelimesinden *standart testler* kelimesine kaydırılacağını belirtmiştir. Yine uzmanların, *“Üçüncü sınıfın üstündeki tüm sınıflarda, kendi sınıf düzeyindeki standart testleri geçemeyen öğrenciler sınıf tekrarı yapmalıdır”* maddesindeki *geçemeyen* ifadesi yerine *başarılı olamayan* ifadesinin kullanılmasının daha uygun olacağını belirtmesi üzerine bu madde *“Üçüncü sınıfın üstündeki tüm sınıflarda, kendi sınıf düzeyindeki standart testlerde başarılı olamayan öğrenciler sınıf tekrarı yapmalıdır”* şeklinde değiştirilmiştir. Uzmanlar arasındaki uyumun yeterli olduğunun belirlenmesi ve uzmanların görüşleri doğrultusunda maddelerin ifade ediliş şekillerinde gerekli değişikliklerin yapılmasıyla üzerinde uzlaşılan nihai bir form elde edilmiştir.

5. Adım: Dilsel Eşdeğerlik Çalışmasına Yönelik İstatistiksel Analizlerin Gerçekleştirilmesi

Elde edilen Türkçe form ile İngilizce form arasındaki tutarlığı belirlemek için öğretmen adaylarının STYİÖ'nün Türkçe formu ile orijinal formundan aldıkları puanlar arasındaki korelasyon hesaplanmış ve iki formdan elde edilen puanlar arasındaki farkın anlamlılığı bağımlı gruplar *t*-testi ile analiz edilmiştir. Bu çalışmalar dilsel eşdeğerlik için yeterli kabul edilmiştir.

Araştırmada STYİÖ'nün dilsel eşdeğerliğine ilişkin çalışmaların tamamlanmasından sonra geçerlik ve güvenirlik ile ilgili çalışmalara geçilmiştir. Geçerlik çalışması kapsamında yapı geçerliği incelenmiştir. Yapı geçerliği için birinci ve ikinci düzey DFA çalışmasına yer verilmiştir. DFA'da sınanan modelin yeterliliğini ortaya koymak üzere pek çok uyum indeksi kullanılmaktadır. Bu çalışmada, yapılan DFA için Ki-Kare Uyum Testi (Chi-Square Goodness, χ^2), karşılaştırmalı uyum indeksi

*Standart Testlere Yönelik İnanç Ölçeği'nin (STYİÖ)
Türkçe Uyarlaması: Geçerlik ve Güvenirlik Çalışması*

(Comparative Fit Index, CFI), iyilik uyum indeksi (Goodness of Fit Index, GFI), düzeltilmiş iyilik uyum indeksi (Adjustment Goodness of Fit Index, AGFI), fazlalık uyum indeksi (Incremental Fit Index, IFI), tahmin hatalarının ortalamasının karekökü (Root Mean Square Error of Approximation, RMSEA), standartlaştırılmış hata kareleri ortalamasının karekökü (Standardized Root Mean Square Residual, SRMR), sıkı normlaştırılmış uyum indeksi (Parsimony Normed Fit Index, PNFI) ve sıkı iyilik uyum indeksi (Parsimony Goodness of Fit Index, PGFI) incelenmiştir. Uyum indekslerine ilişkin dikkate alınması gereken ölçütler hakkında araştırmacılar arasında tam bir uzlaşma bulunmamaktadır (Wetson & Gore, 2006). Bununla birlikte genel olarak; GFI, CFI ve IFI indeksleri için .90 değeri kabul edilebilir uyuma ve .95 değeri mükemmel uyuma işaret etmektedir (Bentler, 1980; Bentler & Bonett, 1980; Marsh, Hau, Artelt, Baumert & Peschar, 2006). AGFI için .85 değeri kabul edilebilir uyumu ve .90 değeri mükemmel uyumu göstermektedir (Schermelleh-Engel & Moosbrugger, 2003). RMSEA için .10 değeri kabul edilebilir uyumu ve .05 değeri mükemmel uyumu ifade etmektedir (Brown & Cudeck, 1993; Byrne & Campbell, 1999). SRMR için ise .05 değeri mükemmel uyum ve .10 değeri kabul edilebilir uyum ölçütü olarak alınmaktadır (Hu & Bentler, 1999; Kline, 2011). PNFI ve PGFI uyum indekslerinin .50'nin üzerinde olması kabul edilebilir uyumu (Meyers, Gamst & Guarino, 2006), .95 ve üzerinde olması ise, mükemmel uyumu göstermektedir (Meydan & Şeşen, 2011).

STYİÖ'nün güvenirliliği; iç tutarlılığa dayalı güvenirlilik belirleme yöntemlerinden biri olan bileşik (yapısal/composite) güvenirlilik yöntemiyle incelenmiştir. Ölçeğin madde ayırt ediciliği ise, düzeltilmiş madde toplam korelasyonu ve %27'lik alt-üst grup karşılaştırmaları ile hesaplanmıştır. Geçerlik ve güvenirlilik analizleri SPSS 20.0 ve LISREL 8.54 programları kullanılarak gerçekleştirilmiştir. Ölçeğin Türkçeye uyarlanması sürecine takip edilen işlemler Tablo 1'de özetlenmiştir.

Tablo 1: STYİÖ'nün Türkçeye Uyarlanması Süreci

	İzin Alınması	-Robert G. Magee ile elektronik posta yoluyla iletişime geçilip ölçeğin uyarlanabileceğine ilişkin gerekli iznin alınması
	Ölçeğin Türkçe Formunun Oluşturulması	-Brislin, Lonner ve Thorndike (1973) tarafından önerilen beş adım takip edilerek ölçeğin Türkçe formunun oluşturulması
	Geçerlik Çalışmaları	-Yapı Geçerliği: Birinci ve İkinci Düzey DFA'nın uygulanması
	Güvenirlilik Çalışması	-Bileşik güvenirlilik katsayılarının hesaplanması
	Madde Analizi	-Düzeltilmiş madde toplam korelasyonu ile %27'lik alt-üst grup karşılaştırmalarına yer verilmesi

Bulgular

Dilsel Eşdeğerlik

Dilsel eş değerklik çalışması kapsamında, Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi İngilizce Öğretmenliği programında öğrenim gören 59 öğretmen adayına ölçeğin önce özgün formu daha sonra Türkçe formu uygulanmıştır. STYİÖ'nün dilsel

eşdeğerlik çalışmasından elde edilen bulgular, Türkçe ve orijinal formda yer alan maddeler arasındaki korelasyonların .49 ile .92 arasında değiştiğini göstermiştir. Türkçe ve orijinal formlardan elde edilen puanlar arasındaki korelasyon; standart testlerin objektifliği hakkındaki inançlar alt ölçeği için .72, standart testlere dayalı olarak alınan kararlar hakkındaki inançlar alt ölçeği için .89 ve standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar alt ölçeği için .93 ve ölçeğin geneli için .86 olarak bulunmuştur. STYİÖ'nün Türkçe ve İngilizce formlarında yer alan maddelerin korelasyonlarına ilişkin bulgular Tablo 2'de gösterilmiştir.

Tablo 2. STYİÖ Maddelere Göre Dilsel Eşdeğerliğe İlişkin Korelasyon Analizi Sonuçları

Madde No	İngilizce Form	Türkçe Form	r
M1	Standardized tests are the best way to objectively measure how much a student knows	Standart testler bir öğrencinin ne kadar bildiğini objektif olarak ölçmenin en iyi yoludur.	.80
M2	Students are too different for a single standardized test to really be useful in measuring their abilities.	Öğrenciler çok farklı özelliklere sahip olduğundan, öğrencilerin yeteneklerini ölçmede standart bir test yararlı olmaz.	.49
M3	A good standardized test can provide a fair (unbiased) indication of the quality of education a student receives in school.	İyi bir standart test, öğrencilerin okulda aldığı eğitimin kalitesi ile ilgili yansız bir gösterge sağlayabilir.	.90
M4	True knowledge is too complex to be measured by a standardized test.	Gerçek bilgi, standart test ile ölçülemeyecek kadar karmaşıktır.	.76
M5	It is impossible for a standardized test to really be unbiased	Standart testlerin gerçekten tarafsız (yansız) olması imkânsızdır.	.92
M6	It is a good idea for states to require all high school students to pass a standardized test or set of standardized tests in order to graduate from high school.	Ülkelerin, liseden mezun olabilmeleri için tüm öğrencilerinden bir veya daha fazla standart testte başarı olmalarını istemeleri iyi bir fikirdir.	.92
M7	Students, in any grade above third grade, who do not pass their grade-level standardized test should have to repeat that grade level.	Üçüncü sınıfın üstündeki tüm sınıflarda, kendi sınıf düzeyindeki standart testlerde başarılı olamayan öğrenciler sınıf tekrarı yapmalıdır.	.80
M8	Schools whose students have the highest scores on standardized tests should receive bonus money.	Standart testlerden en yüksek puanı alan öğrencilerin okulları para ödülü almalıdır.	.91
M9	Teachers whose students score the highest on standardized tests should receive bonus money.	Standart testlerden en yüksek puan alan öğrencilerin öğretmenleri para ödülü almalıdır.	.88

Araştırmada STYİÖ'de yer alan maddelerin Türkçe ve İngilizce formunun dilsel açıdan eş değer olup olmadığını belirlemek için maddeler arasındaki korelasyonları hesaplamanın yanı sıra, Türkçe ve İngilizce formdan alınan puanlar arasındaki farkın anlamlılığı bağımlı gruplar *t*-testi ile analiz edilmiştir. Analize ilişkin bulgular Tablo 3'te sunulmuştur.

*Standart Testlere Yönelik İnanç Ölçeği'nin (STYİÖ)
Türkçe Uyarlaması: Geçerlik ve Güvenirlik Çalışması*

Tablo 3. STYİÖ Maddelere Göre Dilsel Eş Değerliğe İlişkin Bağımlı Gruplar t-testi Sonuçları

Madde No	Formlar	$\bar{X}$	ss	t	Madde No	Formlar	$\bar{X}$	ss	t
M1	Türkçe	2.47	1.12	.55	M6	Türkçe	3.03	1.08	.05
	İngilizce	2.42	1.33			İngilizce	2.92	1.16	
M2	Türkçe	1.55	.90	.27	M7	Türkçe	2.32	1.07	.57
	İngilizce	1.69	.96			İngilizce	2.27	1.05	
M3	Türkçe	2.98	1.00	.41	M8	Türkçe	2.31	1.22	1.00
	İngilizce	2.93	1.08			İngilizce	2.31	1.21	
M4	Türkçe	2.22	1.04	.47	M9	Türkçe	2.20	1.14	.11
	İngilizce	2.29	1.02			İngilizce	2.08	1.10	
M5	Türkçe	2.71	1.15	1.00					
	İngilizce	2.71	1.16						

sd=58, p>.05

Tablo 3 incelendiğinde, STYİÖ'nün Türkçe ve İngilizce formundan alınan puanlar arasındaki farka ilişkin t-değerlerinin .11 ile 1.00 arasında değiştiği ve hiçbir farkın istatistiksel açıdan anlamlı olmadığı belirlenmiştir. Bağımlı gruplar t-testi ile korelasyon analizinden elde edilen bulgular, STYİÖ'nün Türkçe ve İngilizce formunun dilsel açıdan eş değer olduğuna yönelik bir kanıt olarak değerlendirilmiştir.

Yapı Geçerliği

Yapısal eşitlik modeli analizlerinden elde edilen sonuçlar, uygulamanın yapıldığı örneklem ve benzer özellikleri taşıyan gruplar ile sınırlıdır. Farklı özellikler gösteren birden fazla grup söz konusu olduğunda, modelin analizine yönelik olarak her bir gruptan ayrıca veri toplanması ve bu veriler üzerinden modele ilişkin hipotezlerin test edilmesi gerekmektedir (Kline, 2011). Bu nedenle, STYİÖ'nün Türkçe formunun faktör yapısı incelenirken, öğretmenler ve öğretmen adaylarından toplanan veriler ayrı ayrı analize alınmıştır. Bu şekildeki bir uygulama sayesinde özgün formu üniversite öğrencileri üzerinde geliştirilen STYİÖ'nün Türk örneklemindeki öğretmenler ve öğretmen adayları için geçerli sonuçlar verip vermeyeceğinin test edilmesi amaçlanmıştır. Gerek öğretmenlerden gerekse de öğretmen adaylarından toplanan veriler üzerinden gerçekleştirilen yapı geçerliği çalışmalarında, DFA kullanılmıştır. Ölçeğin orijinal formunda elde edilen üç faktörlü yapının Türk örneklemindeki öğretmen ve öğretmen adayları üzerinde doğrulanıp doğrulanmadığı birinci ve ikinci düzey DFA ile incelenmiştir. Aşağıda ilk olarak öğretmenlerden toplanan veriler üzerinden gerçekleştirilen birinci ve ikinci düzey DFA'ya ilişkin bulgular sunulmuştur. Daha sonra, öğretmen adaylarından elde edilen veriler üzerinden yürütülen DFA sonuçlarına yer verilmiştir.

Öğretmenlerden Toplanan Veriler Üzerinden Gerçekleştirilen Yapı Geçerliği Çalışması

Öğretmenlerden toplanan veriler üzerinden gerçekleştirilen birinci düzey DFA'da, minimum χ^2 değerinin ($\chi^2=55.88$, $N=361$, $p=.00$) anlamlı olduğu

görülmüştür. Birinci düzey DFA için elde edilen uyum indeksi değerleri ise, $\chi^2/sd=2.33$, CFI=.94, GFI=.97, AGFI=.94, IFI=.94, RMSEA=.061, SRMR=.055, PNFI=.60 ve PGFI=.52 olarak bulunmuştur. Sınanan modelin yeterliğini ortaya koymak amacıyla incelenen uyum indekslerine ilişkin kabul edilebilir ve mükemmel uyum değerleri ile birinci düzey DFA'dan elde edilen uyum indeksi değerleri ve bu doğrultuda ortaya çıkan sonuçlar Tablo 4'te gösterilmiştir.

Tablo 4. İncelenen Uyum İndekslerine İlişkin Mükemmel ve Kabul Edilebilir Uyum Değerleri ile Birinci Düzey DFA'dan Elde Edilen Uyum İndeksi Değerleri

İncelenen Uyum İndeksleri	Mükemmel Uyum	Kabul Edilebilir Uyum	Birinci Düzey DFA'dan Elde Edilen Uyum İndeksleri	Sonuç
χ^2/sd	$0 \leq \chi^2/sd \leq 2$	$2 \leq \chi^2/sd \leq 3$	2.33	Kabul Edilebilir Uyum
CFI	$.95 \leq CFI \leq 1.00$	$.90 \leq CFI \leq .95$	.95	Mükemmel Uyum
GFI	$.95 \leq GFI \leq 1.00$	$.90 \leq GFI \leq .95$	.97	Mükemmel Uyum
AGFI	$.90 \leq AGFI \leq 1.00$	$.85 \leq AGFI \leq .90$	.94	Mükemmel Uyum
IFI	$.95 \leq IFI \leq 1.00$	$.90 \leq IFI \leq .95$	.95	Mükemmel Uyum
RMSEA	$.00 \leq RMSEA \leq .05$	$.05 \leq RMSEA \leq .10$	.061	Kabul Edilebilir Uyum
SRMR	$.00 \leq SRMR \leq .05$	$.05 \leq SRMR \leq .10$	.050	Mükemmel Uyum
PNFI	$.95 \leq PNFI \leq 1.00$	$.50 \leq PNFI \leq .95$	.61	Kabul Edilebilir Uyum
PGFI	$.95 \leq PGFI \leq 1.00$	$.50 \leq PGFI \leq .95$	.52	Kabul Edilebilir Uyum

Tablo 4'teki uyum indekslerine ilişkin mükemmel ve kabul edilebilir uyum ölçütleri, birinci düzey DFA'dan elde edilen üç faktörlü modelin uyumlu olduğunu ve ölçeğin orijinal formundaki faktör yapısının Türk örneklemindeki öğretmenler üzerinde doğrulandığını göstermektedir. Araştırma grubundaki öğretmenlerden elde edilen veriler kullanılarak gerçekleştirilen birinci düzey DFA'da, STYİÖ'deki her bir madde için hesaplanan *t*-değerleri ise Tablo 5'te sunulmuştur.

Tablo 5. STYİÖ için Birinci Düzey DFA'dan Elde Edilen *t*-testi Değerleri

Madde No	<i>t</i>	Madde No	<i>t</i>	Madde No	<i>t</i>
M1	10.03*	M4	8.45*	M7	7.44*
M2	7.03*	M5	4.64*	M8	13.86*
M3	8.67*	M6	7.29*	M9	12.75*

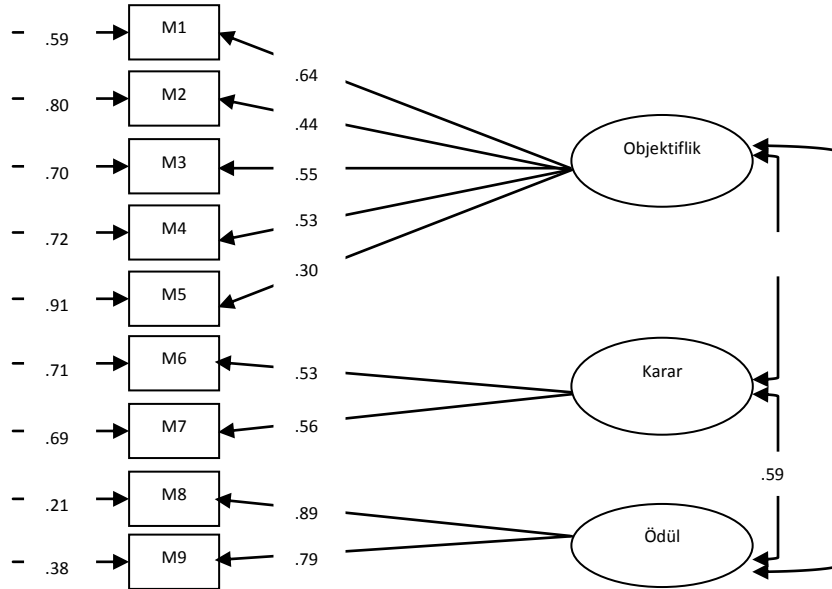
**p* < .01

Tablo 5'teki bulgulara göre, STYİÖ'de yer alan maddelere ilişkin *t*-değerlerinin 4.64 ile 13.86 arasında değiştiği belirlenmiştir. Hesaplanan *t*-değerlerinin 1.96'dan büyük olması .05 düzeyinde; 2.58'den büyük olması ise .01 düzeyinde anlamlı olduğunu göstermektedir (Jöreskog & Sörbom, 1993; Kline, 2011). Buna göre, birinci düzey DFA'da elde edilen tüm *t*-değerlerinin .01 düzeyinde anlamlı olduğu saptanmıştır. Anlamlı olmayan *t*-değerleri, söz konusu *t*-değerlerine ilişkin maddelerin modelden çıkarılması gerektiğine işaret etmekte veya araştırmadaki katılımcı sayısının faktör analizi için yetersiz olduğunun göstergesi olarak değerlendirilmektedir (Byrne, 2010). Dolayısıyla, birinci düzey DFA sonucunda elde edilen *t*-değerleri, araştırma grubundaki öğretmen sayısının faktör

*Standart Testlere Yönelik İnanç Ölçeği'nin (STYİÖ)
Türkçe Uyarlaması: Geçerlik ve Güvenirlik Çalışması*

analizi için yeterli olduğunu doğrulamakta ve modelden çıkarılması gereken madde bulunmadığını ortaya koymaktadır.

Birinci düzey DFA'dan elde edilen üç boyutlu modele ilişkin faktör yükleri Şekil 1'de görülmektedir. Şekil 1'de görülebileceği gibi, standart testlerin objektifliği hakkındaki inançlar (Objektiflik) boyutu için faktör yükleri .30 ile .64 arasındadır. Standart testlere dayalı olarak alınan kararlar hakkındaki inançlar (Karar) boyutunda yer alan maddelerin faktör yükleri .53 ve .56'dır. Standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar (Ödül) boyutuna ait maddelerin faktör yükleri ise .79 ve .89 olarak belirlenmiştir.

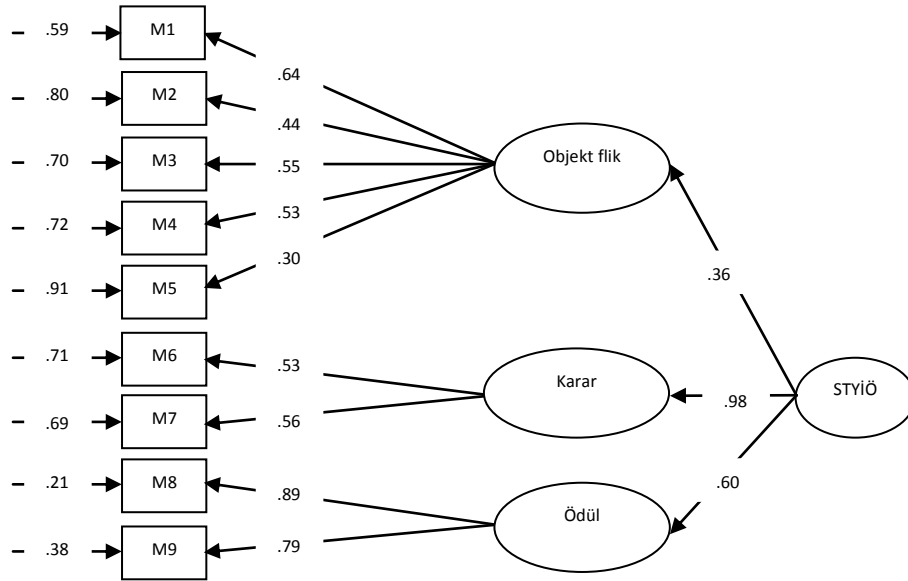


Şekil 1: STYİÖ için Öğretmenlerden Elde Edilen Veriler Üzerinden Uygulanan Birinci Düzey DFA'ya ilişkin Ölçüm Modeli

Ölçeğin orijinal formuna ait üç faktörlü yapının Türk örneklemindeki öğretmenler üzerinde doğrulanıp doğrulanmadığını belirlemek amacıyla gerçekleştirilen birinci düzey DFA'ya ek olarak, standart testlerin objektifliği hakkındaki inançlar, standart testlere dayalı olarak alınan kararlar hakkındaki inançlar ve standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar alt ölçeklerinin bir üst yapı olarak tanımlanan "Standart Testlere Yönelik İnanç" örtük değişkenine ne ölçüde uyum sağladığını belirlemek amacıyla, ikinci düzey DFA uygulanmıştır. Diğer bir deyişle, birinci düzey DFA'da birbirinden göreceli olarak bağımsız ancak birbiriyle ilişkili birer temel bileşen olan standart testlerin objektifliği hakkındaki inançlar, standart testlere dayalı olarak alınan kararlar hakkındaki inançlar ve standart testlere dayalı olarak yapılan ödüllendirmeler

hakkındaki inançlar alt ölçeklerinin daha üst düzey bir yapı olan Standart Testlere Yönelik İnançların birer bileşeni olup olmadığı ikinci düzey DFA ile sınanmıştır.

İkinci düzey DFA’da, minimum χ^2 değerinin ($\chi^2=55.87$, $N=361$, $p=.00$) anlamlı olduğu görülmüştür. İkinci düzey DFA için elde edilen uyum indeksi değerleri ise, $\chi^2/sd=2.33$, CFI=.95, GFI=.97, AGFI=.94, IFI=.95, RMSEA=.061, SRMR=.050, PNFI=.61 ve PGFI=.52 olarak bulunmuştur. Buna göre, birinci ve ikinci düzey DFA’dan elde edilen uyum indekslerinin eş değer olduğu belirlenmiştir. İkinci düzey DFA’dan elde edilen uyum indeksleri, modelin uyumlu olduğunu göstermektedir. Bu bulguya dayanarak, Standart Testlere Yönelik İnanç; standart testlerin objektifliği hakkındaki inançlar, standart testlere dayalı olarak alınan kararlar hakkındaki inançlar ve standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar olarak adlandırılan üç faktörlü bir yapı ile ölçülebileceği söylenebilir. İkinci düzey DFA’da sınanan hiyerarşik modelin yeterli uyum vermesi, ölçeğin tüm maddelerinin toplanabilir olduğuna işaret etmektedir. İkinci düzey DFA’dan elde edilen modele ilişkin faktör yükleri Şekil 2’de görülmektedir.



Şekil 2: STYİÖ için Öğretmenlerden Elde Edilen Veriler Üzerinden Uygulanan İkinci Düzey DFA’ya ilişkin Ölçüm Modeli

Şekil 2’de görülebileceği gibi, standart testlerin objektifliği hakkındaki inançlar alt boyutu için faktör yükleri .30 ile .64 arasında değişmektedir. Standart testlere dayalı olarak alınan kararlar hakkındaki inançlar alt boyutundaki maddelerin faktör yükleri .53 ve .56, standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar alt boyutunda yer alan maddelerin faktör yükleri ise .89 ve .79 olarak belirlenmiştir. İkinci düzey DFA’dan elde edilen uyum

*Standart Testlere Yönelik İnanç Ölçeği'nin (STYİÖ)
Türkçe Uyarlaması: Geçerlik ve Güvenirlik Çalışması*

indeksleri, faktör yükleri, hata varyansı ve tahmin değerlerinin (estimates) birinci düzeyde DFA'da hesaplanan uyum indeksleri ve parametreler ile aynı olduğu belirlenmiştir. Bu sonuç, ikinci düzey ilişkilerin modeldeki uyum indeksi ve parametre değerleri üzerinde bir etki yaratmadığı anlamına gelmektedir.

Öğretmenlerden toplanan veriler üzerinden gerçekleştirilen DFA sonuçları, STYİÖ'nün Türkçe formunun yapı geçerliğinin sağlandığına işaret etmektedir. Bununla birlikte, ölçeğin Türkçe formunun faktör yapısına ilişkin elde edilen bulguların öğretmen adaylarında da geçerli olup olmadığının tespiti için öğretmen adaylarından toplanan veriler üzerinden gerçekleştirilecek bir DFA çalışmasına ihtiyaç duyulmaktadır. Bu kapsamda, 133 öğretmen adayından toplanan veri kullanılarak birinci ve ikinci düzey DFA çalışması yürütülmüştür.

Öğretmen Adaylarından Toplanan Veriler Üzerinden Gerçekleştirilen Yapı Geçerliği Çalışması

Öğretmen adaylarından toplanan veriler üzerinden yürütülen birinci düzey DFA'da, minimum χ^2 değerinin ($\chi^2=40.92$, $N=133$, $p=.00$) anlamlı olduğu belirlenmiştir. Birinci düzey DFA için elde edilen uyum indeksi değerleri ise, $\chi^2/sd=1.71$, CFI=.94, GFI=.94, AGFI=.88, IFI=.94, RMSEA=.073, SRMR=.069, PNFI=.60 ve PGFI=.51 olarak bulunmuştur. Sınanan modelin yeterliğini ortaya koymak amacıyla incelenen uyum indekslerine ilişkin kabul edilebilir ve mükemmel uyum ölçütleri ile birinci düzey DFA'dan elde edilen uyum indeksi değerleri ve bu doğrultuda ortaya çıkan sonuçlar Tablo 6'da gösterilmiştir.

Tablo 6. İncelenen Uyum İndekslerine İlişkin Mükemmel ve Kabul Edilebilir Uyum Değerleri ile Birinci Düzey DFA'dan Elde Edilen Uyum İndeksi Değerleri

İncelenen Uyum İndeksleri	Mükemmel Uyum	Kabul Edilebilir Uyum	Birinci Düzey DFA'dan Elde Edilen Uyum İndeksleri	Sonuç
χ^2/sd	$0 \leq \chi^2/sd \leq 2$	$2 \leq \chi^2/sd \leq 3$	1.71	Mükemmel Uyum
CFI	$.95 \leq CFI \leq 1.00$	$.90 \leq CFI \leq .95$	.94	Kabul Edilebilir Uyum
GFI	$.95 \leq GFI \leq 1.00$	$.90 \leq GFI \leq .95$	.94	Kabul Edilebilir Uyum
AGFI	$.90 \leq AGFI \leq 1.00$	$.85 \leq AGFI \leq .90$	.88	Kabul Edilebilir Uyum
IFI	$.95 \leq IFI \leq 1.00$	$.90 \leq IFI \leq .95$	.94	Kabul Edilebilir Uyum
RMSEA	$.00 \leq RMSEA \leq .05$	$.05 \leq RMSEA \leq .10$	.073	Kabul Edilebilir Uyum
SRMR	$.00 \leq SRMR \leq .05$	$.05 \leq SRMR \leq .10$	.069	Kabul Edilebilir Uyum
PNFI	$.95 \leq PNFI \leq 1.00$	$.50 \leq PNFI \leq .95$	.58	Kabul Edilebilir Uyum
PGFI	$.95 \leq PGFI \leq 1.00$	$.50 \leq PGFI \leq .95$	.50	Kabul Edilebilir Uyum

Tablo 6'daki uyum indekslerine ilişkin mükemmel ve kabul edilebilir uyum ölçütleri, birinci düzey DFA'dan elde edilen üç faktörlü modelin uyumlu olduğunu ve ölçeğin orijinal formundaki faktör yapısının Türk örneklemindeki öğretmen adayları üzerinde doğrulandığını ortaya koymaktadır.

Öğretmen adaylarından toplanan veriler üzerinden gerçekleştirilen birinci düzey DFA'da, STYİÖ'deki her bir madde için hesaplanan t -değerleri Tablo 7'de

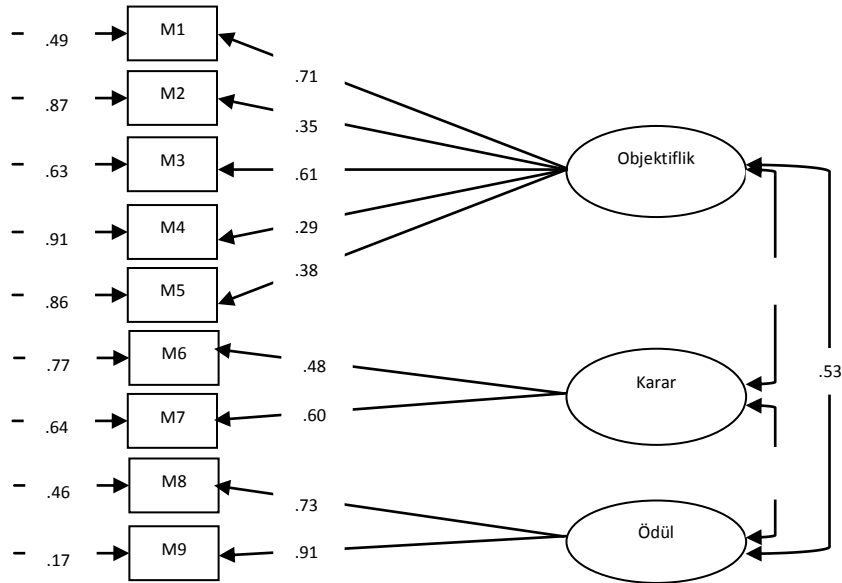
sunulmuştur. Tablo 7’deki bulgulara göre, STYİÖ’de yer alan maddelere ilişkin t değerlerinin 2.89 ile 9.34 arasında değiştiği ve tüm t -değerlerinin istatistiksel olarak anlamlı olduğu saptanmıştır. Buna göre, DFA çalışmasına dâhil edilen öğretmen adayı sayısının faktör analizi için yeterli olduğu ve modelden çıkarılması gereken madde bulunmadığı söylenebilir.

Tablo 7. STYİÖ için Birinci Düzey DFA’dan elde Edilen t -testi Değerleri

Madde No	t	Madde No	t	Madde No	t
M1	7.16*	M4	2.89*	M7	4.50*
M2	3.50*	M5	3.78*	M8	7.79*
M3	6.19*	M6	4.05*	M9	9.34*

$p < .01$

Birinci düzey DFA’dan elde edilen üç boyutlu modele ilişkin faktör yükleri Şekil 3’te görülmektedir. Şekil 3’te görülebileceği gibi, standart testlerin objektifliği hakkındaki inançlar (Objektiflik) boyutu için faktör yükleri .29 ile .71 arasındadır. Standart testlere dayalı olarak alınan kararlar hakkındaki inançlar (Karar) boyutunda yer alan maddelerin faktör yükleri .48 ve .60’dır. Standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar (Ödül) boyutuna ait maddelerin faktör yükleri ise .73 ve .91 olarak belirlenmiştir.



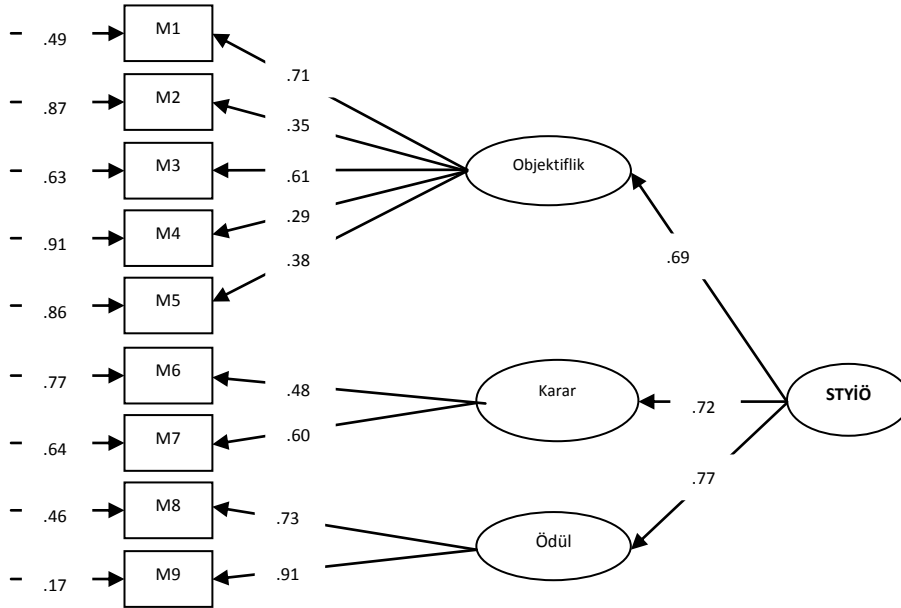
Şekil 3: STYİÖ için Öğretmen Adaylarından Elde Edilen Veriler Üzerinden Uygulanan Birinci Düzey DFA’ya ilişkin Ölçüm Modeli

Ölçeğin orijinal formuna ait üç faktörlü yapının Türk örneklemindeki öğretmen adayları üzerinde doğrulanıp doğrulanmadığını belirlemek amacıyla uygulanan birinci düzey DFA’ya ek olarak, standart testlerin objektifliği hakkındaki

*Standart Testlere Yönelik İnanç Ölçeği'nin (STYİÖ)
Türkçe Uyarlaması: Geçerlik ve Güvenirlik Çalışması*

inançlar, standart testlere dayalı olarak alınan kararlar hakkındaki inançlar ve standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar alt ölçeklerinin daha üst düzey bir yapı olarak tanımlanan *Standart Testlere Yönelik İnanç* örtük değişkenin anlamlı bir yordayıcısı olup olmadığı ikinci düzey DFA ile sınanmıştır.

İkinci düzey DFA'da, minimum χ^2 değerinin ($\chi^2=40.92$, $N=133$, $p=.00$) anlamlı olduğu saptanmıştır. İkinci düzey DFA'da hesaplanan uyum indeksi değerleri ise, $\chi^2/sd=1.71$, CFI=.94, GFI=.94, AGFI=.88, IFI=.94, RMSEA=.073, SRMR=.069, PNFI=.58 ve PGFI=.50 olarak bulunmuştur. Buna göre, birinci ve ikinci düzey DFA'dan elde edilen uyum indeksi değerlerinin eş değer olduğu saptanmıştır. İkinci düzey DFA'dan elde edilen uyum indeksleri, modelin uyumlu olduğunu ve öğretmen adaylarının standart testlere yönelik inançlarının; standart testlerin objektifliği hakkındaki inançlar, standart testlere dayalı olarak alınan kararlar hakkındaki inançlar ve standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar olarak adlandırılan üç faktörlü bir yapı ile ölçülebileceğini göstermektedir. İkinci düzey DFA'dan elde edilen modele ilişkin faktör yükleri Şekil 4'te görülmektedir.



Şekil 4: STYİÖ için Öğretmen Adaylarından Elde Edilen Veriler Üzerinden Hesaplanan İkinci Düzey DFA'ya İlişkin Ölçüm Modeli

Şekil 4'te görülebileceği gibi, standart testlerin objektifliği hakkındaki inançlar alt boyutu için faktör yükleri .29 ile .71 arasında değişmektedir. Standart

testlere dayalı olarak alınan kararlar hakkındaki inançlar alt boyutundaki maddelerin faktör yükleri .48 ve .60, standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar alt boyutunda yer alan maddelerin faktör yükleri ise .73 ve .91 olarak belirlenmiştir. İkinci düzey DFA'dan elde edilen uyum indeksleri, faktör yükleri, hata varyansı ve tahmin değerlerinin birinci düzeyde DFA'da hesaplanan uyum indeksi değerleri ve parametreler ile eş değer olduğu saptanmıştır. Bu durum, ikinci düzey ilişkilerin modeldeki uyum indeksleri ve parametre değerleri üzerinde bir etki yaratmadığını göstermektedir.

Güvenirlilik

STYİÖ'nün güvenirliliği bileşik güvenirlilik yöntemiyle incelenmiştir. Bileşik güvenirlilik, iç tutarlılığa dayalı bir güvenirlilik katsayısı olup DFA'dan elde edilen faktör yükleri ve hata varyansları kullanılarak hesaplanmaktadır (Yang & Green, 2011). STYİÖ'nün güvenirlilik katsayıları, öğretmenler ve öğretmen adaylarından toplanan veriler için ayrı ayrı hesaplanmıştır. Öğretmenlerden toplanan veriler üzerinden hesaplanan bileşik güvenirlilik katsayıları; standart testlerin objektifliği hakkındaki inançlar alt ölçeği için .62, standart testlere dayalı olarak alınan kararlar hakkındaki inançlar alt ölçeği için .41, standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar alt ölçeği için .75 ve ölçeğin geneli için .66 şeklinde bulunmuştur. Öğretmen adaylarından elde edilen veriler kullanılarak hesaplanan güvenirlilik değerleri incelendiğinde, standart testlerin objektifliği hakkındaki inançlar, standart testlere dayalı olarak alınan kararlar hakkındaki inançlar ve standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar alt ölçekleri ile ölçeğin geneli için elde edilen bileşik güvenirlilik katsayıları sırasıyla; .59, .41, .71 ve .77 olarak tespit edilmiştir.

Güvenirlilik katsayısı .70 ve üzerinde olan ölçeklerin güvenilir olduğu kabul edilmektedir (Tezbaşaran, 1997). Ancak, iç tutarlık katsayıları ölçekte yer alan madde sayısından etkilenmeye açık olup, ölçekteki madde sayısı arttıkça yükselme eğilimi gösterdiğinden (Urbina, 2004) madde sayısı az olan ölçeklerde, güvenirlilik katsayısının .50'yi aşması ölçeğin güvenirliliği için yeterli görülmektedir (Raines-Eudy, 2000). Standart testlere dayalı olarak alınan kararlar hakkındaki inançlar alt ölçeği için hesaplanan güvenirlilik katsayısı dışındaki güvenirlilik değerleri bu ölçütü karşılar niteliktedir. Madde sayısı az olan ölçme araçlarında, iç tutarlık ölçütü olarak; Cronbach Alpha ya da bileşik güvenirlilik katsayılarının yanı sıra maddeler arasındaki korelasyon katsayılarından da yararlanabilmektedir. Madde sayısı az olan ölçekler için maddeler arasındaki korelasyon katsayılarının .20'nin üzerinde olması, ölçeğin güvenirliliği için yeterli kabul edilmektedir (Briggs & Cheek, 1986). Nitekim, STYİÖ'nün özgün formunda da ölçeğin geneli ile standart testlerin objektifliği hakkındaki inançlar alt boyutunun güvenirliliği Cronbach Alpha iç tutarlık katsayısı ile incelenirken; yalnızca 2'şer madde içeren standart testlere dayalı olarak alınan kararlar ve standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar alt boyutlarının güvenirliliği bu boyutlarda yer alan maddeler arasındaki korelasyon katsayıları hesaplanarak tespit edilmiştir. Bu araştırmada, standart

*Standart Testlere Yönelik İnanç Ölçeği'nin (STYİÖ)
Türkçe Uyarlaması: Geçerlik ve Güvenirlik Çalışması*

testlere dayalı olarak alınan kararlar hakkındaki inançlar alt ölçeğinde yer alan maddeler arasındaki korelasyon katsayıları; öğretmenlerden elde edilen veriler için $r=.30$, $p<.001$ ve öğretmen adaylarından elde edilen veriler için $r=.29$, $p<.001$ olarak hesaplanmıştır. Buna göre, standart testlere dayalı olarak alınan kararlar alt ölçeğine ilişkin güvenilirlik değerlerinin de kabul edilebilir sınırlar içerisinde yer aldığı söylenebilir.

Madde Analizi

STYİÖ'de yer alan maddelerin toplam puanı yordama gücünü belirlemek ve ayırt edicilik düzeylerini saptamak amacıyla madde analizi yapılmıştır. Bu kapsamda düzeltilmiş madde toplam korelasyonu hesaplanmış ve %27'lik alt-üst grup karşılaştırmalarına yer verilmiştir. Madde toplam korelasyonunun hesaplanmasında Pearson Momentler Çarpımı Korelasyonu kullanılmıştır. %27'lik alt üst grup karşılaştırmalarında ise bağımsız gruplar t -testinden yararlanılmıştır. Diğer analizlerde olduğu gibi, madde analizinde de öğretmenlerden ve öğretmen adaylarından elde edilen veriler ayrı ayrı analize alınmıştır. Öğretmenlerden toplanan veriler üzerinden gerçekleştirilen madde analizi sonuçları Tablo 8'de; öğretmen adaylarından toplanan veriler üzerinden uygulanan madde analizi sonuçları ise Tablo 9'da sunulmuştur.

Tablo 8. STYİÖ'ye İlişkin Öğretmenlerden Toplanan Veriler Üzerinden Gerçekleştirilen Madde Analizi Sonuçları

Faktörler	Madde No	Madde Toplam Korelasyonu	Gruplar	Ortalama	Standart Sapma	t	
Objektiflik	M1	.44	Üst _{%27}	3.57	.79	10.82**	sd=169 **p<.001
			Alt _{%27}	2.22	.83		
	M2	.35	Üst _{%27}	3.20	1.12	13.68**	
			Alt _{%27}	1.33	.52		
	M3	.36	Üst _{%27}	3.75	.78	11.68**	
			Alt _{%27}	2.27	.87		
Karar	M4	.30	Üst _{%27}	3.56	.94	16.28**	sd=144 **p<.001
			Alt _{%27}	1.52	.66		
	M5	.36	Üst _{%27}	3.58	.85	9.45**	
			Alt _{%27}	2.27	.96		
	M6	.30	Üst _{%27}	4.19	.69	19.88**	
			Alt _{%27}	1.81	.69		
Ödül	M7	.70	Üst _{%27}	4.33	.66	25.68**	sd=176 **p<.001
			Alt _{%27}	1.58	.64		
	M8		Üst _{%27}	4.05	.80	29.91**	
			Alt _{%27}	1.15	.36		
M9	Üst _{%27}		4.34	.60	42.63**		
	Alt _{%27}		1.13	.34			

Öğretmenlerden toplanan veriler üzerinden gerçekleştirilen madde analizine ilişkin bulgular incelendiğinde, %27'lik alt ve üst grupların madde puanlarındaki farklara ait t -değerlerinin standart testlerin objektifliği hakkındaki inançlar alt ölçeği için 9.45 ile 16.28 ($sd=169$, $p<.001$) arasında değiştiği görülmektedir. Standart testlere dayalı olarak alınan kararlar hakkındaki inançlar alt boyutundaki maddeler

için hesaplanan t -değerleri 19.88 ile 25.68 ($sd=144$, $p<.001$) olarak belirlenmiştir. Standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar alt boyutunda bulunan maddelere ilişkin t -değerleri ise, 29.91 ve 42.63 ($sd=176$, $p<.001$) olarak hesaplanmıştır. Alt ve üst grup arasındaki farklara ilişkin t -değerlerinin anlamlı olması maddenin ayırt ediciliği için bir kanıt olarak değerlendirilmektedir (Erkuş, 2012). Dolayısıyla, alt-üst grup karşılaştırmaları sonucunda elde edilen t -değerleri ölçekte yer alan maddelerin ayırt edici olduğuna işaret etmektedir. Tablo 8'e göre ayrıca, madde toplam korelasyonuna ilişkin sonuçlar standart testlerin objektifliği hakkındaki inançlar alt ölçeği için .30 ile .44 arasında değişmektedir. Standart testlere dayalı olarak alınan kararlar hakkındaki inançlar alt boyutunda yer alan maddeler arasındaki korelasyon .30 ve standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar alt boyutunda bulunan maddeler arasındaki korelasyon ise .70 olarak hesaplanmıştır. Madde toplam korelasyonunun yorumlanmasında, değeri .30 ve üzerinde olan maddelerin ölçülecek özelliği ayırt etme açısından yeterli olduğu kabul edilmektedir (Büyüköztürk, 2010; Erkuş, 2012; Field, 2009). Madde analizinden elde edilen bulgular; %27'lik alt-üst grup karşılaştırmaları ve madde toplam korelasyonuna ilişkin ölçütler doğrultusunda değerlendirildiğinde, STYİÖ'de yer alan yer alan maddelerin tamamının ayırt edici olduğu söylenebilir.

Tablo 9. STYİÖ'ye İlişkin Öğretmenler Adaylarından Toplanan Veriler Üzerinden Gerçekleştirilen Madde Analizi Sonuçları

Faktörler	Madde No	Madde Toplam Korelasyonu	Gruplar	Ortalama	Standart Sapma	t	
Objektiflik	M1	.44	Üst _{%27}	3.74	.95	10.34**	$sd=67$ $p<.001$
			Alt _{%27}	1.68	.68		
	M2	.35	Üst _{%27}	2.63	1.35	5.95**	
			Alt _{%27}	1.21	.41		
	M3	.36	Üst _{%27}	3.80	.93	7.96**	
			Alt _{%27}	2.15	.78		
Karar	M4	.30	Üst _{%27}	3.06	1.03	6.92**	$sd=77$ $p<.001$
			Alt _{%27}	1.59	.70		
	M5	.36	Üst _{%27}	3.60	.88	8.34**	
			Alt _{%27}	1.85	.86		
	M6		Üst _{%27}	3.93	.76	11.44**	
		.29	Alt _{%27}	2.00	.73		
Ödül	M7		Üst _{%27}	3.61	.90	14.41**	$sd=83$ $p<.001$
			Alt _{%27}	1.34	.48		
	M8		Üst _{%27}	3.79	1.04	14.10**	
		.67	Alt _{%27}	1.03	.19		
	M9		Üst _{%27}	3.52	.91	14.79**	$sd=83$ $p<.001$
			Alt _{%27}	1.00	.00		

Öğretmen adaylarından elde edilen verilere göre, %27'lik alt ve üst grupların madde puanlarındaki farklara ilişkin t -değerlerinin standart testlerin objektifliği hakkındaki inançlar alt ölçeği için 5.95 ile 10.34 ($sd=67$, $p<.001$) arasında değiştiği saptanmıştır. Standart testlere dayalı olarak alınan kararlar hakkındaki inançlar alt boyutundaki maddeler için hesaplanan t -değerleri 11.44 ve 14.41 ($sd=77$, $p<.001$)

*Standart Testlere Yönelik İnanç Ölçeği'nin (STYİÖ)
Türkçe Uyarlaması: Geçerlik ve Güvenirlik Çalışması*

olarak bulunmuştur. Standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar alt boyutunda bulunan maddelere ilişkin t değerleri ise, 14.10 ve 14.79 ($sd=83$, $p<.001$) olarak saptanmıştır. Alt-üst grup arasındaki farklara ilişkin t -değerlerinin anlamlı olması maddenin ayırt ediciliği için bir kanıt olarak değerlendirilmektedir (Erkuş, 2012). Tablo 9'a göre ayrıca, madde toplam korelasyonuna ilişkin sonuçlar; standart testlerin objektifliği hakkındaki inançlar alt ölçeği için .30 ile .44 arasında değişmektedir. Standart testlere dayalı olarak alınan kararlar hakkındaki inançlar alt boyutunda yer alan maddeler arasındaki korelasyon .29 ve standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar alt boyutunda bulunan maddeler arasındaki korelasyon ise .67 olarak tespit edilmiştir. Bu bulgulardan hareketle, STYİÖ'de yer alan maddelerin tamamının öğretmen adayları için de ayırt edici sonuçlar verdiği söylenebilir.

STYİÖ'den Alınan Puanların Değerlendirilmesi

STYİÖ'de 9 madde bulunmaktadır. Ölçekte *Kesinlikle Katılıyorum (5)* → *Kesinlikle Katılmıyorum (1)* şeklinde 5'li likert tipi bir derecelendirme kullanılmıştır. Ölçek; standart testlerin objektifliği hakkındaki inançlar, standart testlere dayalı olarak alınan kararlar hakkındaki inançlar ve standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar olmak üzere üç boyutlu bir yapıya sahiptir. Standart testlerin objektifliği hakkındaki inançlar alt boyutunda 5 madde bulunduğundan, bu boyuttan alınabilecek en düşük puan 5 en yüksek puan ise 25'dir. Bu boyuttan yer alan maddelerin 2'si (1 ve 3 numaralı maddeler) standart testlere yönelik inanca ilişkin olumlu maddelerdir. Kalan 3 madde (2, 4 ve 5 numaralı maddeler) ise olumsuz olup tersten puanlanmaktadır. Standart testlere dayalı olarak alınan kararlar hakkındaki inançlar alt boyutunda 2 madde bulunmaktadır. Dolayısıyla bu boyuttan alınabilecek en düşük puan 2, en yüksek puan ise 10'dur. Bu boyuttaki maddeler standart testlere yönelik inanca ilişkin olumlu maddelerdir. Bu nedenle, bu boyutta tersten puanlanan madde bulunmamaktadır. Benzer şekilde, standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar alt boyutunda 2 madde bulunduğundan bu boyuttan alınabilecek en düşük puan 2, en yüksek puan ise 10'dur. Bu boyuttaki maddelerin her ikisi de standart testlere yönelik inanca ilişkin olumlu maddelerdir. Dolayısıyla bu boyutta da tersten puanlanan madde bulunmamaktadır.

STYİÖ'nün hem birinci düzey hem de ikinci düzey DFA'da yeterli uyum indeksleri vermesi; ölçeğin alt boyutlarından alınan puanlar üzerinden işlem yapılabileceği gibi ölçekten standart testlere yönelik inanca ilişkin toplam bir puan da elde edilebileceğini ortaya koymaktadır. STYİÖ'nün alt boyutlarından ve ölçeğin genelinden alınan puanların yükselmesi, standart testlere yönelik yüksek inancı göstermektedir.

Tartışma ve Sonuç

Bu çalışmada, Magee ve Jones (2012) tarafından geliştirilen STYİÖ'nün Türkçe formunun geçerlik ve güvenilirliği incelenmiştir. STYİÖ'nün Türkçeye uyarlanması sürecinde ilk olarak ölçeğin İngilizce ve Türkçe formlarının dilsel açıdan eşdeğer olup olmadığı araştırılmıştır. Bu amaçla, İngilizce ve Türkçe form arasındaki korelasyon hesaplanmış ve orijinal ile çeviri formundan alınan puanlar arasındaki farkın anlamlılığı bağımlı gruplar *t*-testi kullanılarak analiz edilmiştir. Dilsel eşdeğerlik çalışmasından elde edilen bulgular; Türkçe ve İngilizce form arasındaki korelasyonun oldukça yüksek olduğunu ortaya koymuş, özgün ve çeviri formdan alınan puanlar arasındaki farkın ölçekte yer alan maddelerin hiç birinde anlamlı olmadığını göstermiştir. Bu sonuçlar, ölçek maddelerinin Türkçe çevirisinin İngilizce formda yer alan orijinal maddelerle benzeştiğini ve Türkçe form ile İngilizce formun dilsel açıdan eş değer olduğunu göstermektedir.

STYİÖ'nün yapı geçerliğini incelemek ve ölçeğin orijinal formundaki üç faktörlü yapının Türk örneklemindeki öğretmen ve öğretmen adayları üzerinde doğrulanıp doğrulanmadığını ortaya koymak amacıyla DFA uygulanmıştır. Öğretmenler ve öğretmen adaylarından toplanan veriler üzerinden uygulanan birinci ve ikinci düzey DFA'dan elde edilen uyum indeksi değerleri; STYİÖ'ye ilişkin üç faktörlü modelin uyumunun yeterli olduğunu göstermiştir. Buna göre, STYİÖ'nün orijinal formunda tespit edilen üç faktörlü yapının ölçeğin Türkçe formu için de geçerli olduğu söylenebilir. STYİÖ'nün gerek öğretmenlerden gerekse de öğretmen adaylarından toplanan veriler kullanılarak yapılan analizlerde kabul edilebilir uyum verdiği dikkate alındığında; hem öğretmenler hem de öğretmen adayları için ölçeğin yapı geçerliğinin sağlandığı söylenebilir.

STYİÖ'nün güvenilirliğini belirlemeye yönelik olarak bileşik güvenilirlik yöntemi kullanılmıştır. Ölçeğin geneli, standart testlerin objektifliği hakkındaki inançlar ve standart testlere dayalı olarak yapılan ödüllendirmeler hakkındaki inançlar alt ölçekleri için hesaplanan bileşik güvenilirlik katsayılarının yeterli olduğu belirlenmiştir. Standart testlere dayalı olarak alınan kararlar hakkındaki inançlar alt ölçeği için hesaplanan bileşik güvenilirlik katsayısının ise düşük olduğu tespit edilmiştir. Bununla birlikte, madde sayısı az olan ölçeklerde maddeler arasındaki korelasyon katsayısının .20'nin üzerinde olmasının ölçeğin güvenilirliğine işaret ettiği bilinmektedir (Briggs & Cheek, 1986) Standart testlere dayalı olarak alınan kararlar hakkındaki inançlar ölçeğinde yer alan maddeler arasındaki korelasyon katsayısının bu ölçütün üzerinde olduğu göz önüne alındığında, bu boyuta ilişkin güvenilirlik değerlerinin kabul edilebilir sınırlar içerisinde yer aldığı sonucuna ulaşılmıştır. STYİÖ'nün orijinal formunda hesaplanan güvenilirlik değerleri ile bu çalışmadan elde edilen katsayıları karşılaştırıldığında; ölçeğin orijinal formunda hesaplanan güvenilirlik katsayılarının Türkçe formundan elde edilen güvenilirlik katsayılarına kıyasla daha yüksek olduğu belirlenmiştir. Güvenirliğin, ölçme aracına ait bir özellik olmayıp; araştırmadan elde edilen ölçümlere ilişkin bir özellik olduğu göz önünde

*Standart Testlere Yönelik İnanç Ölçeği'nin (STYİÖ)
Türkçe Uyarlaması: Geçerlik ve Güvenirlik Çalışması*

bulundurulduğunda, değişik örneklemeler üzerinde yürütülen araştırmalarda farklı güvenirlilik katsayılarının elde edilmesi beklenen bir sonuç olarak değerlendirilebilir.

STYİÖ'de yer alan maddelerin ayırt edicilik düzeylerini saptamak amacıyla düzeltilmiş madde toplam korelasyonu hesaplanmış ve %27'lik alt ve üst grup karşılaştırmalarına yer verilmiştir. Madde analizinden elde edilen bulgulara göre, gerek öğretmenler gerekse de öğretmen adayları için ölçekte yer alan maddelerin tümünün yeterli düzeyde ayırt edici olduğu belirlenmiştir. Araştırmadan geçerlik, güvenirlilik ve madde analizi ilgili elde edilen bulgular, STYİÖ'nün öğretmenlerin ve öğretmen adaylarının standart testler hakkındaki inançlarını ölçmek amacıyla kullanılabilecek geçerli ve güvenilir bir araç olduğunu göstermektedir. Sonuç olarak; Magee ve Jones (2012) tarafından geliştirilen STYİÖ'yü Türk kültürüne kazandırmayı amaçlayan bu çalışma ile öğretmenlerin ve öğretmen adaylarının standart testler hakkındaki inançlarının ölçülmesinde kullanılabilecek yeterli psikometrik özelliklere sahip bir ölçme aracına ulaşıldığı söylenebilir.

Öneriler

Alanyazın incelendiğinde, Türk kültüründe epistemolojik inanç (Deryakulu & Büyüköztürk, 2002; Deryakulu & Hazır Bıkmaz, 2003), öğrenmeye yönelik inanç (Bay vd., 2012), zekâya yönelik inanç (İlhan & Çetin, 2013) ve katılımcı değerlendirmeye yönelik inanç (İlhan, Çetin & Bars, 2013) gibi bireylerin öğrenme-öğretme sürecine ilişkin farklı inançlarını ölçmek amacıyla kullanılabilecek ölçme araçlarının bulunduğu görülmektedir. Ancak Türkçe alanyazında, bireylerin standart testler hakkındaki inançlarını belirlemede kullanılabilecek bir ölçme aracına rastlanmamıştır. Dolayısıyla, STYİÖ'nün Türkçeye uyarlandığı bu çalışmanın önemli olduğu düşünülmekte ve alanyazındaki bu boşluğu dolduracağına inanılmaktadır. Öte yandan, bu araştırmanın birtakım sınırlılıkları bulunmakta ve bu sınırlılıkların ileri araştırmalar ile aşılabileceğine inanılmaktadır.

Öncelikle bu çalışmada, STYİÖ'nün yapı geçerliğine yönelik işlemler DFA ile sınırlı tutulmuştur. STYİÖ'nün yapı geçerliğine ilişkin ek kanıt elde etmek için ileri araştırmalarda STYİÖ'nün yakınsak (convergent) ve ıraksak (divergent) geçerliğinin incelenmesi yerinde olacaktır. Araştırmada, STYİÖ'nün geçerlik çalışmalarına yönelik bir diğer sınırlılık, geçerlik çalışmaları kapsamında uyum geçerliğine yer verilmemiş olmasıdır. Araştırmaya ilişkin bu sınırlılığın aşılabilmesi için standart testlere yönelik inanç ile ilişkili olabilecek yapıları değerlendiren, geçerlik ve güvenirliliği kanıtlanmış ölçme araçları ile STYİÖ arasındaki ilişkilerin incelenmesi çalışmalara ihtiyaç duyulmaktadır. Bu ihtiyaca cevap verebilecek araştırmaların yapılması, STYİÖ'nün uyum geçerliğinin ortaya konulması açısından oldukça önemlidir. Son olarak, bu araştırmada, STYİÖ'nün güvenirliliği yalnızca bileşik güvenirlilik yöntemiyle incelenmiştir. İleri araştırmalarda STYİÖ'nün test tekrar test güvenirliliğinin incelenmesi, ölçeğin zamana karşı değişmezliğinin tespit edilebilmesi açısından önem arz etmektedir.

Kaynaklar

Abd-El-Fattah, S.M., & Yates, G.C.R. (2006, November) *Implicit theory of intelligence scale: Testing for factorial invariance and mean structure*. Paper presented at the Australian Association for Research in Education Conference, Adelaide, South Australia.

Airasian, P.W. (2005). *Classroom assessment: Concepts and applications*. NY: McGraw Hill.

Ajzen, I. (2005). *Attitudes, personality and behavior*. New York: Open University Press.

Aksu Ataç, B. (2012). Foreign language teachers' attitude toward authentic assessment in language teaching. *The Journal of Language and Linguistic Studies*, 8 (2), 7-19.

Başol Göçmen, G. (2004, Temmuz). *Otantik değerlendirme nedir ve nasıl yapılır?* XIII. Ulusal Eğitim Bilimleri Kurultayı'nda sunulmuş sözlü bildiri. Malatya: İnönü Üniversitesi.

Bay, E., İlhan, M., Aydın, Z., Kinay, İ., Özyurt, M., Kuzu, S. vd. (2013). Öğrenmeye yönelik inanç ölçeği: Geçerlik ve güvenirlik çalışması. *Uluslararası İnsan Bilimleri Dergisi*, 9(2), 1008-1030.

Beaton, D.E., Bombardier, C., Guillemin, F., & Ferraz, M.B. (2000). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine*, 25(24), 3186-3191.

Bentler, P.M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, 31, 419-456.

Bentler, P.M., & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.

Brew, C., & Riley, P. (2011). Measuring student support for participative assessment and related strategies: Development and testing of the beliefs about participative assessment instrument (BAPAI). *Australian Journal of Teacher Education*, 36(4), 32-52.

Boud, D. (1992). The use of self-assessment schedules in negotiated learning. *Studies in Higher Education*, 17(2), 185-200.

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment and Evaluation in Higher Education*, 31(4), 399- 413.

*Standart Testlere Yönelik İnanç Ölçeği'nin (STYİÖ)
Türkçe Uyarlaması: Geçerlik ve Güvenirlik Çalışması*

Briggs, S.R., & Cheek, J.M. (1986). The Role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, 54, 106-148.

Brislin, R.W., Lonner, W. J., & Thorndike, R.M. (1973). *Cross-cultural research methods*. New York, John Wiley - SonsPub.

Brown, G.T.L., & Hirschfeld, G.H.F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice*, 15(1), 3-17.

Brown, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In: K. Bollen & J. Long, (Eds), *Testing Structural Equation Models* (pp. 136-162). London: Sage Publications.

Büyüköztürk, Ş. (2010). *Sosyal bilimler için veri analizi el kitabı*. Ankara: Pegem Akademi Yayınları.

Byrne, B. (2010). *Structural equation modeling with AMOS: Basic concepts, applications and programming*. New York, NY: Taylor and Francis Group.

Byrne, B., & Campbell, T.L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-Cultural Psychology*, 30(5), 555-574.

Cronbach, L.J. (1970). *Essentials of psychological testing*. New York: Harper and Row.

Deryakulu, D. (2006). Epistemolojik inançlar, Y. Kuzgun ve D. Deryakulu, (Ed.), *Eğitimde Bireysel Farklılıklar içinde* (261-289). Ankara: Nobel Yayın Dağıtım.

Deryakulu, D., & Büyüköztürk, Ş. (2002). Epistemolojik inanç ölçeğinin geçerlik ve güvenirlik çalışması. *Eğitim Araştırmaları Dergisi*, 8, 1-15.

Deryakulu, D., & Büyüköztürk, Ş. (2005). Epistemolojik inanç ölçeğinin faktör yapısının yeniden incelenmesi: Cinsiyet ve öğrenim görülen program türüne göre epistemolojik inançların karşılaştırılması. *Eğitim Araştırmaları Dergisi*, 18, 57-70.

Deryakulu, D., & Hazır Bıkmaz, F. (2003). Bilimsel epistemolojik inançlar ölçeğinin geçerlik ve güvenirlik çalışması. *Eğitim Bilimleri ve Uygulama*, 2(4), 243-257.

Dikli, S. (2003). Assessment at a distance: Traditional vs. alternative assessments. *The Turkish Online Journal of Educational Technology-TOJET*, 2(3), 1303-6521.

Mustafa İLHAN, Bayram ÇETİN, İsmail KİNAY

Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme*. Ankara: Pegem Akademi Yayınları.

Falchikov, N. (1995). Peer feedback marking developing peer assessment. *Innovations in Education and Training International*, 32, 175-187.

Field, A. (2009). *Discovering statistics using SPSS*. London: SAGE Publications Ltd.

Geisinger, K.F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychol Assess*, 6(4), 304-312.

Graham, M., Milanowski, A., & Miller, J. (2012). *Measuring and promoting inter-rater agreement of teacher and principal performance ratings*. Washington, DC: Center for Educator Compensation Reform. 02 Ocak 2014 tarihinde http://cecr.ed.gov/pdfs/Inter_Rater.pdf adresinden alınmıştır.

Gronlund, N.E., & Linn, R.L. (1990). *Measurement and evaluation in teaching*. New York: Macmillan.

Gullo, D.F. (2005). *Understanding assessment and evaluation in early childhood education*. New York: Teachers College Press.

Gültekin, S. (2012). Performansa dayanlı değerlendirme. N. Çıkrıkçı Demirtaşlı, (Ed.), *Eğitimde Ölçme ve Değerlendirme* içinde (s. 251-280). Ankara: Elhan Yayınları.

Higgins, S., & Moseley, D. (2001). Teachers' thinking about information and communications technology and learning: Beliefs and outcomes. *Teacher Development*, 5, 191-210.

Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structural analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.

İlhan, M., & Çetin, B. (2013). Örtük zekâ teorisi ölçeğinin Türkçe uyarlaması: Geçerlik ve Güvenirlik çalışması. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi (EFMED)*, 7(1), 191-221.

İlhan, M., Çetin, B., & Bars, M. (2013). Katılımcı değerlendirmeye yönelik inanç ölçeği'nin (KDYİÖ) Türkçe uyarlaması: Geçerlik ve güvenirlik çalışması. *Journal of European Education*, 3(2), 17-35.

Jöreskog, K.G., & Sörbom, D. (1993). *LISREL 8: User's guide*. Chicago: Scientific Software.

*Standart Testlere Yönelik İnanç Ölçeği'nin (STYİÖ)
Türkçe Uyarlaması: Geçerlik ve Güvenirlik Çalışması*

Karet, N., & Hubbell, E.R. (2003). *Authentic assessment*. 09 Ocak 2014 tarihinde <http://www.elizrosshubbell.com/portfolio/authenticassessment.pdf> adresinden alınmıştır.

Kline, R.B. (2011). *Principles and practice of structural equation modeling*. New York: The Guilford Press.

Koç, N. (1985). Standart başarı testlerinin, bir eğitim sisteminde verilen çeşitli kararlardaki yeri ve önemi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 17(1-2). Ankara: 159-172.

Koretz, D.M. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.

Krows, A.J. (1999). *Preservice teachers' belief systems and attitudes toward mathematics in the context of a progressive elementary teacher preparation program*. Unpublished Doctoral Dissertations. The University of Oklohama. Norman, Oklohoma.

Landis, J., & Knoch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.

Luft, J.A., & Roehrig, G.H. (2007). Capturing science teachers' epistemological beliefs: The development of the teacher beliefs interview. *Electronic Journal of Science Education*, 11(2), 38-63.

Mansour, N. (2009). Science teachers' beliefs and practices: Issues, Implications and research agenda. *International Journal of Environmental & Science Education*, 4(1), 25-48.

Magee, R.G., & Jones, B.D. (2012). An instrument to assess beliefs about standardized testing: Measuring the influence of epistemology on the endorsement of standardized. *Australian Journal of Educational & Developmental Psychology*, 12, 71-82.

Marsh, H.W., Hau, K.T., Artelt, C., Baumert, J. & Peschar, J.L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing*, 6(4), 311-360.

Marstocn, D., Deno, S., & Tindal, G. (1983). *A comparison of standardized achievement tests and direct measurement techniques in measuring pupil progress* (Res. Rep. No. 126). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities (ERIC Document Reproduction Service No. ED236198).

Mustafa İLHAN, Bayram ÇETİN, İsmail KİNAY

Meisels, S.J., Xue, Y., Bickel, D.D., Nicholson, J., & Burnett, S.A. (2001). Parental reactions to authentic performance assessment. *Educational Assessment*, 7(1), 61-85.

Meydan, C.H. & Şeşen, H. (2011). *Yapısal eşitlik modellemesi ve AMOS uygulamaları*. Ankara: Detay Yayıncılık.

Meyers, L.S., Gamst, G., & Guarino, A.J. (2006). *Applied multivariate research: Design and interpretation*. London: SAGE Publications.

Nichols, J.O., & Nichols, K.W. (2005). *A road map for improvement of student learning and support services through services through assessment*. New York: Agathon.

Olswang, L.B., & Bain, B.A. (1988). Assessment of language in developmentally disabled infants and preschoolers. In T.D. Wachs and R. Sheehan (Ed.). *Assessment of young developmentally disabled children. Perspectives in developmental psychology* (pp.285-320). New York: Plenum Press.

Özgüven, İ.E. (2011). *Psikolojik testler*. Ankara: PDREM Yayınları.

Pajares, F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307-332.

Paul, M., Owence, C., Chrispen, C., & Alexander, C.R. (2013). Peer assessment in higher education: The roadmap for developing employability skills in potential job seekers. *International J. Educational & Research*, 1(2), 62-69.

Pope, N. (2001). An examination of the use of peer rating for formative assessment in the context of the theory of consumption values, *Assessment & Evaluation in Higher Education*, 26(3), 235-246.

Putnam, J. (2005). *Advantage test prep, grade 7: High-interest skill building for home and school*. Huntington Beach: Creative Teaching Press.

Raines-Eudy, R. (2000). Using structural equation modeling to test for differential reliability and validity: An empirical demonstration. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(1), 124-141.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, 82(3), 498-504.

Schermelleh-Engel, K., Moosbrugger, H., & Muller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, 8(2), 23-74.

Shipley, K., & McAfee, J. (2009). *Assessment in speech-language pathology: A resource manual*. Clifton Park, NY: Delmar Cengage Learning.

*Standart Testlere Yönelik İnanç Ölçeği'nin (STYİÖ)
Türkçe Uyarlaması: Geçerlik ve Güvenirlik Çalışması*

Sikka, A., Nath, J.L., & Cohen, M.D. (2007). Practicing teachers beliefs and uses of assessment. *International Journal of Case Method Research & Application*, XIX(3), 239-253.

Silver, E.A., & Kenney, P.A. (1995): Sources of assessment information for instructional guidance in mathematics. En T.A. Romberg (Ed.), *Reform in School Mathematics and Authentic Assessment*. Albany: State University of New York Press.

Stein-Rubin, C., & Fabus, R. (2012). *A guide to clinical assessment and professional report writing in speech-language pathology*. Clifton Park, NY: Delmar/Cengage Learning.

Tezbaşaran, A. (1997). *Likert tipi ölçek hazırlama kılavuzu*. Ankara: Türk Psikologlar Derneği.

Urbina, S. (2004). *Essentials of psychological testing*. New Jersey: John Wiley & Sons. Inc.

Van den Berg, I., Admiraal, W., & Pilot, A. (2006). Design principles and outcomes of peer assessment in higher education. *Studies in Higher Education*, 31(3), 341-356.

Yang, Y., & Green, S.B. (2011). Coefficient Alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment*, 29(4) 377-392.

Walvoord, B.E. (2010). *Assessment clear and simple: A practical guide for institutions, departments and general education*. San Francisco: John Wiley and Sons Inc.

Weston, R., & Gore, P.A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist*, 34(5), 719-751.

Woolley, S.L., Benjamin, W. J.J., & Woolley, A.W. (2004). Construct validity of a self-report measure of teacher beliefs related to constructivist and traditional approaches to teaching and learning. *Educational and Psychological Measurement*, 64, 319-331.