Ankara University, Journal of Faculty of Educational Sciences, year: 2007, vol: 40, no: 1, 1-16

The Adaptation of Psychological Scales

K. Zülfikar DENİZ*

ABSTRACT. The exchange of knowledge and information is rapidly progressing and the intercultural interaction is gradually developing in these days. The psychological scales developed for a culture can also be used in different ones. Therefore, it must be answered that how appropriate in another culture the scale is?, how strong the psychometrics of scale (reliability and validity are)? and how accurately the scale reflects the psychological properties of new culture?. Adaptation is to translate the scale into another language and to search answers to these questions above.

The aim of this study is to depict the adaptation process of psychological measurement scales, the projects of TIMMS, TIMMS-R, OECD/PISA, the reason of adaptation instead of developing a measurement scale, International Test Commision (ITC), ITC activities and test adaptation guide.

Key Words: Adaptation, Psychological Scale, Instrument, Development, Construction steps, ITC, TIMSS-R, OECD/PISA

SUMMARY

Today, globalization and intercultural interaction are very important in economic, social and cultural fields. Moreover, these will be important in the future more than today, because interaction among people is gradually

* Research Assisstant, Ankara University Faculty of Educational Sciences, Department of Measurement and Evaluation, **zlfkrdnz@yahoo.com, deniz@education.ankara.edu.tr**

increasing in time. As a result of these circumstances most of the scientific researches appeal to the intercultural studies. Similarly, a lot of psychological studies are the comparison of the psychological properties of people in different cultures or different countries. While comparing psychological characteristics of people in different cultures using a psychological scale developed for a culture can be required in different languages and cultures. For instance, the most common international tests are The Programme for International Student Assessment (PISA) of OECD and The Trends in International Mathematics and Science Study (it is generally known as Third International Mathematics and Science Studies or briefly TIMSS). "The Programme for International Student Assessment (PISA) is an internationally standardised assessment that was jointly developed by participating countries and administered to 15-year-olds in schools. The survey was implemented in **43 countries** in the first assessment in 2000, in **41 countries** in the second assessment in 2003 and at least **58 countries** will participate in the third assessment in 2006. Tests are typically administered to between 4,500 and 10,000 students in each country" (pisa.oecd, 2006). Equally, TIMSS is an internationally standardised assessment that 46 countries joined in 2003.

To carry out these and similar standardised assessments, a scale must be adapted to a lot of cultures and languages. Adaptation is to translate the instrument and find out the answer that how appropriate in a new culture the scale is, psychometrics of the scale (reliability, validity) are how adequate, the scale is how accurately introduce to psychological properties in new culture.

Briefly, in this study, the importance of the adaptation has been depicted and the answer has been searched for the question: "why adaptation instead of developing?". Adaptation is useful and necessary, when cross-cultural studies are especially carried out or when we want to use, in our study, a valid and reliable scale developed in another language. The aim of this research is to find out how we can adapt a scale, which steps must be followed in an adaptation process. etc… Then, International Test Comission (ITC) and its activities have been mentioned.

NOTE: This is an extended summary of the following article originally written in Turkish.

Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, yıl: 2007, cilt: 40, sayı: 1, 1-16

Psikolojik Ölçme Aracı Uyarlama

K. Zülfikar DENİZ*

ÖZ. Son yıllarda, bilgi hızla çoğalmakta ve kültürler arası etkileşim giderek artmaktadır. Bu etkileşimin bir yansıması da bir kültür için geliştirilen psikolojik ölçeklerin başka kültürlerde de kullanılmasıdır. Ölçeklerin farklı kültürlerde kullanılabilmesi için ölçeğin kullanılacağı yeni kültüre ne derece uygun olduğu, yeni kültürdeki psikometrik özelliklerinin (güvenirlik, geçerlik) ne derece yeterli olduğu, ölçeğin yeni kültürdeki psikolojik özellikleri ne derece iyi yansıttığı gibi bir takım sorulara yeterli yanıtların verilmesi gerekir. Ölçeğin çevirisinin farklı bir kültüre yapıldığı ve bu sorulara cevap arandığında, ölçek uyarlama sürecinden söz edilebilir.

Bu çalışmada ölçme aracı uyarlama, TIMSS, TIMSS-R, OECD/PISA projeleri, "neden ölçek geliştirme yerine uyarlama?" sorusu, uyarlamanın aşamaları, Uluslararası Test Komisyonu (International Test Commission(ITC)) ve faaliyetleri, ITC'nin test uyarlama rehberine değinilmektedir. Çalışmanın, çeşetli alanlarda ölçek geliştirmeyi veya uyarlamayı planlayan araştırmacılara yol gösterici olacağı beklenmektedir.

Anahtar Sözcükler: Uyarlama, psikolojik ölçek, geliştirme aşamaları, ITC, TIMSS-R, OECD/PISA

* Araş.Gör, Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Anabilim Dalı, **zlfkrdnz@yahoo.com, deniz@education.ankara.edu.tr**

GİRİŞ

Günümüzde belli bir kültür için hazırlanmış olan bir psikolojik ölçme aracı farklı kültür ve dillere çevrilerek de kullanılmaktadır. Bir ölçeğin yalnızca başka dile çevrilip kullanılması yerine o ölçekle ilgili temel psikometrik işlemlerin (geçerlik, güvenirlik) de yapılması süreci ölçeğin başka dil ve kültürlere uyarlanması olarak bilinmektedir.

Cronbach (1990) bir yetenek testini alan bazı kültürlerin ya da alt kültürlerin o yetenek ile ilgili yapabileceklerinin en iyisini ortaya koyamayabileceklerini belirtmektedir. Çeşitli kültürlerin benzer şeylere farklı tepkiler verdiğini belirten Cronbach (1990) ZuÑi çocuklarının yarıştan çok birlikte iş yapmalarının o kültür için geçerli olduğunu "tahtada en hızlı problem çözme becerisini ölçmek için çıkarılan çocukların en son kişi bitirene kadar bitirdiklerini göstermemeleri" örneğiyle vurgulamaktadır. Oysa bizim okullarımızda bu şekilde yapılan bir sınama durumunda bütün öğrenciler çabuk bitirmek için ellerinden gelen çabayı göstereceklerdir. Bu bizim öğrencilerimizin seçme sınavları sebebiyle sürekli yarış halinde olması yanında, anne-babaların çocuklarını yetiştirme tarzlarının farklılığından kaynaklandığı da söylenebilir.

Amerika Birleşik Devletleri'nde yapılan pek çok çalışmada siyahların genel yetenek ortalamasının beyazlardan yaklaşık bir standart sapma aşağıda olduğu ortaya konulmuştur. Okul çağındaki çocuklarda ise altıncı sınıf civarına kadar siyah çocukların beyaz çocuklardan üstün olduğu görülmektedir. Bu yüzden puanları yorumlarken dikkatli olmak gerekir (Cronbach, 1990). Puanların yorumlanmasında her grubun özellikleri dikkate alınmalıdır. Eğer grupların özellikleri bilinmiyorsa karşılaştırmalarda önce bu değerler ortaya konulmalıdır.

Bazı bilim adamları zihinsel olarak zayıf olan siyah ve kızıl derili çocukların sistematik olarak hak ettiklerinden daha az puan aldıklarını belirtmektedirler. Bazıları da ortalamalar arasında gözlenen farkların alt kültürlerde doğuştan geldiğine inanmaktadırlar (Cronbach, 1990). Burada kast edilen durum, doğuştan gelen niteliklerin çevreyle etkileşimi sonunda ortaya çıkan farklılaşmadır.

Sireci ve Berberoğlu (2000) test ve envanterlerin başka kültürlere çevrilmesini, farklı kültürlerdeki insanların başarı, tutum, kişilik ve diğer psikolojik özelliklerinin karşılaştırılması için yaygın bir yöntem olarak görmektedirler. Ancak bir ölçek veya envanterin bir dilden başka bir dile çevrilmesinin, çevrilen dildeki ölçekle asıl dildeki ölçeğin eşit olacağının garantisi olmadığını vurgulamaktadırlar.

Dünya gittikçe daha küçük hale gelmektedir. Uluslar ekonomik, sağlık ya da bilimsel alanlarda gittikçe daha fazla işbirliği içine girmektedirler. Bu

uluslararası alandaki birliktelik başka kültürleri araştıran ve uluslararası araştırmalar yapan sosyal bilimciler için çok güzel fırsatlar sağlamaktadır. Bu alanda mevcut araştırma dergilerinde ve konferanslarda yapılan aktivite yönünde büyük bir patlama yaşanmaktadır (Sireci ve Berberoğlu, 2000).

Hambleton ve Jong (2003) da çok dilli eğitimsel ve psikolojik testlerin her geçen gün arttığını belirtmektedirler. Buna örnek olarak, Third and Fourth International Mathematics and Science Studies (TIMSS and TIMSS-R)'in kırktan fazla ülkede uygulandığını ve otuzdan fazla dile çevrildiğini belirtmektedirler. Kendi ülkelerinin 1995 ve 1999 yılında katıldığı TIMSS ve 2000 yılında katıldığı PISA projelerindeki başarı düzeylerini karşılaştıran Geske ve Kangro (2002) TIMSS maddelerinin çalışmasının, katılımcı ülkelerden gelen uzmanlar tarafından görüşülerek, mümkün olduğunca kısa, kolay okunabilir ve gereksiz ayrıntılardan arındırılmış olarak oluşturulduğunu belirtmektedirler. Ülkelerin eğitim sistemlerini değerlendirmek için 15 yaşındaki öğrencilerin okuma, matematik ve fen bilimleri alanında test edildiği The Programme for International Student Assessment (PISA) of OECD (OECD/PISA) projesinin ilki 2000 yılında(43 ülkede), ikincisi ise 2003 yılında(41 ülkede) olmak üzere 32 dilde ve 40'tan fazla ülkede uygulaması yapılmıştır (Hambleton ve Jong, 2003; OECD, 2004) . Son olarak 2006 yılında 58 ülkede uygulaması yapılmıştır(pisa.oecd, 2006). Türkiye 2000 yılındaki ilk PISA projesine katılmamış, 2003 yılındaki ikinci PISA projesi içinde yer almıştır (OECD, 2004). Türkiye 2006 yılında yapılan son çalışmada da yer almıştır. Geske ve Kangro (2002) PISA projesinde testin içeriğinin TIMSS'te olduğu gibi katılımcı ülkelerden uzmanların görüşlerinin alınmadığını belirtmektedirler. Spielberger'in Trait-State Anxiety Inventory ve Wechsler Intelligence Scale for Children ölçekleri her biri ellinin üzerinde dile çevrilmiş olan tanınmış kişilik ve zeka ölçekleridir. Ayrıca pek çok başarı testi, kimlik sınavı, kişilik ölçeği ve yaşam standardı ölçeği gibi ölçekler de en az 10 kültürde uyarlaması yapılmış ölçeklerdir (Hambleton ve Jong, 2003).

Bu liste günümüzde gittikçe artmaktadır. Buna ek olarak gelecekte daha fazla ölçeğin çeviri ve adaptasyonunun yapılması beklenmektedir. Çünkü:

- Ölçme araçlarının uluslararası değişimleri yaygınlaşmaktadır;
- Novell ve Microsoft gibi firmalar tarafından yapılan kişilik testleri pek çok ülkede kullanılmaktadır;
- Başka kültürlerin psikolojisi ve uluslararası karşılaştırma çalışmaları gittikçe artmaktadır.

Öncelikle, çevrilecek ya da uyarlanacak testlerin ve araçların asıl sürümlerinden büyük olasılıkla farklı olacakları bilinmelidir. Amaç bu

farklılıkların anlamsal, psikometrik, dil ve psikolojik özellik açılarından kabul edilebilir düzeyde olmalarını sağlamaktır (Hambleton ve Jong, 2003).

Neden Uyarlama?

"Neden ölçek geliştirmek yerine uyarlama yapmak?" sorusuna Hambleton ve Patsula (1999) literatürde gösterilen sebepleri 5 maddede özetlemektedirler:

1. Çoğunlukla bir test uyarlamak ikinci kültürde yeni bir test geliştirmekten daha ucuzdur ve daha hızlıdır.
2. Yapılacak testin amacı kültürel ya da ulusal değerlendirme yapmak olduğunda, uyarlanmış bir test, ikinci kültürde denk bir test geliştirmenin en etkili yoludur.
3. İkinci bir kültürde test geliştirmek için uzmanlık bilgisi yetersiz olabilir.
4. Asıl test iyi bilinen bir test olduğunda, o testin uyarlamasının vereceği güven duygusu yeni geliştirilecek olan bir teste duyulacak güvenden daha fazla olacaktır.
5. Bir testin çok kültürlü sürümlerinden çıkarılan sonuçlar testi alan adaylar için çoğunlukla doğru olacaktır.

Bu maddelerde uyarlama çalışması yapmanın yeni bir test geliştirmekten daha kolay olduğu belirtilmektedir. Ancak uyarlama çalışması yapmanın da bazı sıkıntıları vardır.

Bunlardan en önemlisi, ölçeğin asıl kültürüne ve uyarlanacak kültüre hakim olan ve bu iki kültürün karşılaştırmasını yapabilecek düzeyde bilgiye sahip, eğer çeviri/uyarlama farklı dile yapılacaksa, her iki dile de hakim olan, konu alanı uzmanı ve ölçek uyarlama konusunda çalışabilecek uzman bulma sıkıntısıdır. Eğer bu belirtilen özelliklerin hepsine sahip olan bir uzman grubu varsa bu, uyarlama çalışması için oldukça iyi bir durumdur.

Uyarlama konusunda karşılaşılabilecek diğer bir sıkıntı ise yukarıda belirtilen özelliklere sahip olan grubun birlikte çalışma zorunluluğudur. Şöyle ki: Eğer belirtilen bu özelliklerin tamamına sahip olan bir uzman grubu yoksa bu özelliklerden her birisi/birkaçı için farklı uzmanlar belirlenecektir. Dolayısıyla bu belirlenen uzman grubundaki kişiler kendi alanı dışındaki konularda bilgiye sahip olmadıkları için, bu grubun hep birlikte çalışması (yani bir komisyon gibi) gerekecektir. Bu da bu kişiler için uygun zamanlama yapılması güçlüğünü ortaya çıkaracaktır.

Uyarlamada karşılaşılabilecek diğer bir sıkıntı ise, uyarlanan testin psikometrik özelliklerinin sağlanamaması sorunu olabilir. Şöyle ki: Eğer uyarlanacak ölçek aslında yüksek bir güvenirlik ve geçerliğe sahipse, ki

böyle olması istenen durumdur, uyarlanan ölçeğin de bu değerlere yakın değerler vermesi beklenir. Yüksek bir güvenirlik ve geçerlik elde etme, temelde yeni geliştirilen ölçekler için de mevcuttur. Ancak uyarlamadaki farklılık, uyarlanan ölçekte iyi bir güvenirlik ve/veya geçerlik elde edilse bile, bu değer asıl ölçeğin güvenirlik ve/veya geçerlik değerlerinden farklı olabilir. Bu da beraberinde iyi psikometrik özelliklere sahip olmasına rağmen asıl ölçekten daha düşük değerlerde psikometrik özelliklere sahip bir uyarlanmış ölçek getirecektir.

Bu sıkıntılar aşılamayacak veya kesinlikle karşımıza çıkacak olan sıkıntılar değildir. Ancak uyarlama çalışması belli bir plan ve sistem içinde yapılmazsa, karşılaşılması olası durumlar dikkate alınmazsa ya da olası sıkıntıları giderici önlemler alınmazsa uyarlanan ölçeğimizin gücü ve etkisi zayıflayacaktır.

Uyarlamanın Aşamaları

Hambleton ve Patsula (1999) ITC'nin hazırladığı rehberden ve diğer pek çok deneysel çalışmalardan elde edilen bilgilerden yola çıkarak ve özellikle Geisinger (1994)'in çalışmasından esinlendiklerini belirttikleri uyarlama aşamalarını maddeler halinde belirtmekte ve bu maddelere eklemeler, maddelerden çıkartmalar ya da yeni sınıflamaların olabileceğini de vurgulamaktadırlar. Bu çalışmada bu bilgiler doğrultusunda, asıl kaynakta yer alan sınıflamaya araştırmacı tarafından eklemeler ve maddelerin yerlerinde değişiklikler getirilmiştir. Daha sonra yapılacak çalışmalarda da amaca bağlı olarak bu maddelere eklemeler veya çıkartmalar yapılabileceği söylenebilir.

1. **Yeni bir test geliştirmenin mi, yoksa var olan testi uyarlamanın mı daha kullanışlı olacağına karar verilmelidir** (Hambleton ve Patsula, 1999).

Yeni bir test geliştirmenin ve uyarlama çalışması yapmanın avantaj ve dezavantajları dikkate alınarak ve amacımıza uygun bir aracın bizim kültürümüz veya başka kültürler için önceden yapılmış olup olmadığını dikkate alarak bu karar verilmelidir.

Uyarlamaya karar verildiyse, ikinci aşamadan devam edilmelidir.

2. **Uyarlama çalışmasına karar verildiğinde yapılacak olan ilk iş izin almaktır.** İzin alınmadan çalışmaların sürdürülmesinin etik olmayacağı dikkate alınacak olursa, çalışmaya başlamadan önce izin alınması zorunluluğu açıkça ortaya çıkmaktadır.

   Uyarlama için izin üç farklı amaç için alınabilir:

- Eğitsel bir çalışma yapmak amacıyla,
- Araştırma yapmak amacıyla,

- Ölçeğin pazarlamasını yapmak amacıyla

Bu üç farklı amaç için isteyeceğimiz izinlerin cevabı da farklı olabilir.

Karşı taraf bizden uyarlama çalışmasını yapmak için yeterliliğimizin olduğuna dair belge talep edebilir (diploma, daha önceden yapılmış çalışmalar, referans mektubu). Bu belgenin yanısıra araştırma veya pazarlama amaçlı yapacağımız çalışmalarda ölçeğin telif haklarına sahip olan kişi ya da kurum bizden telif ücreti talep edebilir. Bu yüzden ölçeği hangi amaçla uyarlayacağımız izin yazısında belirtilmelidir.

3. **Çalışılacak olan kültürlerdeki ve dil gruplarındaki ölçülen özelliğe ait yapının varlığı ve eşitliği sağlanmalıdır**. Araştırma yapılacak olan farklı kültürlerde aynı psikolojik yapının varlığı araştırılmalıdır; eğer aynı yapı yoksa yapılacak düzeltmelerle incelenen her kültürde ortak bir yapı sağlanmalıdır. Bu mümkün değilse çalışma durdurulmalıdır. Harkness (1998) yapıların eşitliğinin sağlanması çalışması sırasında bize yardımcı olacak olan soruları şu şekilde belirtiyor (Hambleton ve Patsula, 1999):

- Bir araştırmacı ilgilenilen her kültürde ölçülecek yapının varlığını araştırdı mı?
- Bu yapının karşılaştırılmasının yapılabilirliği mümkün mü?
- Yapılan kültürler arası karşılaştırma anlamlı mı?
- Ölçülecek olan yapı karşılaştırılacak tüm kültürlerde aynı anlama mı gelmektedir?

4. **İyi çevirmenler seçilmelidir.** Bir uyarlama çalışmasında ihmal edilen temel noktalardan biri genellikle bu kısımdır. Bunu yapabilmek için:

Çevirmen seçerken, her iki dilde de akıcı konuşan, çalışma yapılan kültürlere aşina, test yapısı ve ölçülen yapı hakkında bir miktar bilgiye sahip olmalarına dikkat edilmelidir.

Bazı araştırmacılar çevirmenleri bir panel veya komite şeklinde toplayarak ayrı ayrı yapılan çeviriye göre daha iyi sonuç almaktadırlar. Bu şekilde yapılan çeviriler ayrı ayrı yapılan çevirilerden daha doğru olmaktadır.

5. **Test çevrilip uyarlamaya geçilmelidir.** Çeviri konusunda yapılacak uygulama bir dilden diğer dile çevirme ya da bir dilden diğerine çeviri yapıldıktan sonra tekrar asıl diline çevirme şeklinde olmalıdır. Özellikle asıl dilden uyarlanacak kültüre yapılan ilk çeviri konusunda 4. maddede belirtilen özelliklerde çevirmen seçilmelidir. Geri çeviride ise çevirmenler ilk çeviriyi yapanlar kadar uzman olmayabilirler, ama hedef kültürlerdeki kelimelere aşina olmalılar.

6. **Ölçeğin uyarlanmış hali gözden geçirilmeli ve gerekiyorsa değişiklikler yapılmalıdır.** Çeviriyi yapan çevirmenler ölçekte yapılması gereken değişiklikler varsa onları yapmalılar. Bu konuda Geisinger (1994) çevirmenlerin birlikte çalışmasını veya bir toplantı düzenlenmesini önermektedir. Çevirmenler çok iyi de olsa, uyarlanan ölçek üzerinde düzeltmeler yapmalarına rağmen tanımlayamadıkları maddeler olabilir. Hambleton, Slater ve Yu tarafından NAEP (National Assessment of Educational Progress)'in Çince'ye adaptasyonu çalışmasında, NAEP'in test maddeleri çevirmenler tarafından tanımlanamamıştır. Çünkü 8. sınıfta okuyan Çinli öğrencilerin bazı kavramlara aşina olmadıkları görülmüştür.

7. **Uyarlanan test deneme grubunda uygulanmalıdır.** Uyarlanan ve düzeltmeler yapılan test, psikometrik özelliklerinin incelenmesine geçilmeden önce pilot gruba uygulanmalı ve ölçekle ilgili yapılması gereken başka düzeltmeler olup olmadığına bakılmalıdır. Bu pilot grubun özellikleri ölçeğin esas uygulamasının yapılacağı grubun özelliklerini taşımalıdır. Buradan elde edilecek sonuçlar da asıl gruptaki sonuçlarla karşılaştırılmalıdır.

8. **Uyarlama aşamasındaki ölçek daha büyük bir grupta uygulanmalıdır.** Bütün test uyarlama sürecinin en önemli aşamalarından birisidir. Güvenirlik ve geçerlik analizleri için, hedef evreni temsil edecek kadar büyük bir örneklem grubuna ölçek uygulanmalıdır. Buna ek olarak da faktör analiziyle yapı kontrol edilmelidir.

9. **Asıl ve uyarlanan kültürlerdeki test puanlarını karşılaştırmak için uygun bir istatistiksel yöntem seçilmelidir**. Bu adım, eğer kültürler arası karşılaştırma yapılacaksa veya kaynak kültürde normlar ya da performans standartları varsa gereklidir. Bu aşama uygulanacaksa 8. adımla birlikte hangi yöntemin kullanılacağına karar verilmelidir. En çok kullanılan üç desen:

a) İki dilli grup deseni,

b) Tek dilli karşılaştırmalı grup deseni,

c) Tek dilli grup deseni.

Bu üç desen de popüler olmasına karşın en kolay olanı üçüncüsüdür.

10. **Eğer kültürler arası karşılaştırma yapılacaksa, testin dil sürümlerinin eşitliğinden emin olunmalıdır.** Bu, uyarlama sürecinde en fazla ön plana çıkarılması gereken adımdır. Ölçek evrenden alınan örnekleme uygulanır ve istatistiksel analizle madde yanlılığına yönelik çalışmalar yapılır (genellikle, differential item functioning DIF çalışması). Her grupta farklı fonksiyona sahip maddeler varsa, bu maddelerin her iki

grupta da aynı fonksiyona sahip olması için yeniden yazım veya yeniden çeviri, yeniden uygulama ve yeniden analiz yapılmalıdır.

11. **Uygun bir yöntemle geçerlik çalışması yapılmalıdır.** Karşılaştırılacak gruplar ne olursa olsun uyarlanan ölçekte geçerlik ve güvenirlik çalışması yapılmalıdır. Geçerlik için yapı geçerliğini sağlamaya yönelik olan faktör analizi yöntemi ya da deneysel veya ilişkisel çalışmalar (yordama veya uygunluk geçerlikleri gibi) yapılmalıdır. Bu adım bundan önceki 3 adımla birleştirilebilir.

12. **Uyarlanan testin kullanıcıları için süreç ve hazırlamayı içeren bir el kitabı hazırlanmalıdır.** 1. adımdan 11. adıma kadar olan süreci içeren bir el kitabı hazırlanmalıdır. Ayrıca el kitabında, testin uygulanması ve sonuçların yorumlanmasıyla ilgili özel bilgiler de bulunmalıdır. Bu çok önemli bir adımdır ama genellikle göz ardı edilir.

13. **Kullanıcılar eğitilmelidir.** Testi uygulayacak ya da kullanacak olan bireylere eğitim verilmelidir.

14. **Uyarlanan test takip edilmelidir.** Genellikle ölçekler bir kere kullanım için uyarlanırlar. Popüler zeka, yeterlilik, yetenek ve kişilik sınavları gibi bazı testler, uyarlanan kültürde sürekli kullanmak için de uyarlaması yapılmış olabilir. Uyarlama yapan araştırmacılar ölçeğin sürekliliği için devamlı uyanık olmalı ve ölçeği takip etmeliler. Yeniden geçerlik ve güvenirlik çalışmaları yapmalılar.

Görüldüğü gibi uyarlaması yapılacak olan çalışmanın niteliğine ve amacına bağlı olarak adımlardan bazıları süreçten çıkarılabilir. Gerekiyorsa çalışmamızı güçlendirici eklemeler de yapılabilir. Ancak belirtilen bazı temel noktalarda değişiklikler yapılmaması gerekmektedir. Aksi halde uyarlama amacına hizmet etmeyecektir.

Uluslararası Test Komisyonu (International Test Commission (ITC)):

ITC, ulusal psikoloji kurumları, test komisyonları, yayıncıları ve diğer etkili test ve değerlendirme politikaları sunan ve geliştiren, değerlendiren, eğitimsel ve psikolojik araçları kullanan bir kurumdur (Intestcom, 2004). 1960'ların ortalarında Jean Cardinet tarafından kurulmaya yönelik çalışmalar başlatılan ITC, resmi olarak 1976'da (the Congress of the International Union of Psychological Sciences (IUPsyS) sırasında kurulmuştur (Oakland, Poortinga, Schlegel ve Hambleton, 2001). Temmuz 2000 itibariyle 27 tam üye (ulusal profesyonel psikoloji kurumları) ve 58 de ITC'ye bağlı örgütü vardır (Diğer test komisyonları, yayıncıları ve test etmeyle ilgili araştırma organizasyonları). Mevcut üyeleri, Doğu ve Batı Avrupa ülkeleri, Kuzey Amerika, Orta ve Uzak Doğu'daki bazı ülkeler, Güney Amerika ve Afrika ülkeleridir.

ITC, üyeleri arasında bilgi değişimi, psikolojik testlerin ve diğer psikolojik teşhis testlerinin yapısı, dağıtımı ve kullanımıyla ilgili sorunlarda ortaklarıyla hareket eden bir kurumdur. Bunları yapmak için ITC, uluslararası toplantılar ve görüşmeler yapmakta, kendisiyle ve diğer bilimsel çalışmaların anlamlandırılmasıyla ilgili çalışmaların basımını teşvik etmektedir. Testlerin bilimsel ve etik kurallar çerçevesinde kullanılmasıyla ilgili araştırma projelerinde uluslararası işbirliğini teşvik etmektedir (Intestcom, 2004).

ITC, 1992'de testlerin ve psikolojik araçların çevirisi ve uyarlaması ve farklı dil ve/veya kültürlerdeki puanların karşılaştırılabilirliğini sağlamak için bir rehber hazırlama projesi başlatmıştır. Pek çok kurum ITC'nin projesine destek olmuştur:

- the European Association of Psychological Assessment,
- the European Test Publishers Group,
- the International Association for Cross-Cultural Psychology,
- the International Association for the Evaluation of Educational Achievement,
- the International Language Testing Association,
- the International Union of Psychological Science.

Bu kurumlardan 12 temsilci (Türkiye'den Prof. Dr. Işık Savaşır) birkaç yıl içinde 22 maddelik bir rehber oluşturdular. Sonra bu maddeler alanda test edildi. Daha sonra, bu rehberin ulusal psikoloji topluluklarına, test basan kurumlara ve araştırmacılara dağıtılması için ITC tarafından onaylandı (Oakland, Poortınga, Schlegel, ve Hambleton, 2001).

ITC'nin Test Uyarlama Rehberi

Pratik açıdan ITC'nin rehberinde iki ana noktada durulmaktadır:

- **Çeviri/uyarlama testlerin yapılması**

Bir veya daha fazla dilde ve/veya ulusal bağlamda bazı ülkelerde kullanıma uygun, özel bir dilde geliştirilen testler ve araçlardan söz edilmektedir. Genellikle, bu tür durumlarda çeviri/uyarlamanın amacı bir test veya aracın aslıyla psikometrik özelliklerini karşılaştırabilmektir. Sözel olmayan testlerde bile, yalnızca uygulama ve puanları yorumlama için değil, aynı zamanda kültürel eğilimlerden kaçınmak için de testin net materyallerine ihtiyaç vardır.

- **Uluslararası kullanım için yeni araçlar geliştirilmesi.**

Uluslararası karşılaştırma için uygun olacak kavramlardan testler ve araçlar geliştirmeyi içermektedir. Bunun avantajı farklı kültürler için yeniden çevirme süreci ve psikometrik özelliklerin saptanması süreci gibi süreçlere gerek kalmamasıdır. Bunun sıkıntılı yanı ise çok çeşitli versiyonlarda ölçeği geliştirmenin zorluğu ve bu sürece dahil olacak insanların sayısının fazlalığıdır.

Bu rehberin maddeleri dört ana başlık altında ve toplam 22 maddede şu şekilde düzenlenmiştir (Hambleton ve Jong, 2003):

1. Genel Durum:

G.D. 1: Çalışmanın temel amaçları için konuyla ilgili olmayan ya da önemli olmayan kültürel farklılıkların etkileri mümkün olduğunca en düşük seviyeye indirilmelidir.

G.D. 2: Çalışılan gruptaki test veya araç tarafından ölçülen yapının binişme miktarı ölçülmelidir.

2. Test Geliştirme ve Uyarlama:

G.A. 1: Test geliştirenler ya da yayınlayanlar, uyarlama sürecinin dildeki bütün farklılıklara ve aracın uyarlandığı gruplar arasındaki kültürel farklılıklara dikkat etmelidir.

G.A. 2: Test geliştirenler ya da yayınlayanlar, hem ölçme aracının uyarlanacağı bütün dil grupları ve kültürler için uygun bir el kitabının hazırlanması hem de yönerge, kural ve maddelerde de söz konusu kültürün kendi dilinin kullanılmasını sağlamalıdır.

G.A. 3: Test geliştirenler ya da yayınlayanlar, seçecekleri test tekniklerini, madde formatlarını, test durumlarını ve işlemlerini herkes için uygun olabilecek düzeyde seçmelidir.

G.A. 4: Test geliştirenler ya da yayınlayanlar, madde içerikleri ve uyarıcı materyalleri herkes için aşina öğelerden seçmelidir.

G.A. 5: Test geliştirenler ya da yayınlayanlar, test sürecinin doğru işlemesi için hem dil hem de psikolojik yönden sistematik bir yol izlemeli ve bütün diller için ortak noktalar bulmalıdır.

G. A. 6: Test geliştirenler ya da yayınlayanlar, testlerde ve araçlarda toplanacak verilerin farklı dillerde istatistiksel analize uygun olacak maddelerden oluşmasını sağlamalıdır.

G.A. 7: Test geliştirenler ya da yayınlayanlar, ölçme aracının farklı versiyonlarının eşitliğini sağlamak için ve bütün evren için yetersiz olabileceği durumları tanımlamak için uygun istatistiksel teknikleri uygulamalıdır.

G.A. 8: Test geliştirenler ya da yayınlayanlar, uyarlanacak hedef evrende geçerliğin değerlendirilmesi konusunda bilgi sağlamalıdır.

G.A. 9: Test geliştirenler ya da yayınlayanlar, testin uyarlandığı bütün gruplar için ölçek maddelerinin eşitliğini istatistiksel olarak sağlamalıdır.

G.A. 10: Genel bir ölçek hazırlarken ya da farklı kültürler karşılaştırılırken, bütün uyarlanacak gruplar arasında denk olmayan maddeler kullanılmamalıdır. Ancak, her kültürün puanları ayrı ayrı rapor edilirken kapsam geçerliğini artırmak için bu sorular kullanılabilir.

3. Uygulama

U.1: Test geliştirenler ya da uygulayanlar, olası problemleri kestirebilmek ve bu problemleri çözmek için uygun materyal ve öğretimleri hazırlamak için deneme yapmalılar.

U.2: Testi uygulayanlar puanların geçerliğinin makul düzeyde olmasını sağlamalıdır. Geçerliğin uyarıcı materyalleriyle, uygulama süreçleriyle ve cevaplama şekliyle ilgili pek çok faktörden etkilendiğini dikkate almalıdır.

U.3: Testin uygulama koşulları farklı kültürlerde de mümkün olduğunca aynı olmalıdır.

U.4: Kaynak ve hedef kültürlerde uygulama koşullarında meydana gelebilecek beklenmedik durumları minimize etmek için, uygulayıcılara test uygulama eğitimleri verilmelidir.

U.5: Testin el kitabı yeni kültürün kapsamında incelenmesi gereken bütün uygulama şekillerine göre özelleştirilmelidir.

U.6: Uygulayıcı işgüzar olmamalı ve uygulayıcı-uygulama etkileşimi minimum düzeyde olmalıdır. Uygulama için el kitabında belirtilen kesin kurallara uyulmalıdır.

4. Raporlaştırma ve Puan Yorumları

R.1: Bir test ya da araç başka bir kültürde kullanılmak üzere uyarlandığında, eşitliklerin kanıtlanması yanında değişikliklerin de raporlaştırılması sağlanmalıdır.

R.2: Test ya da aracın uygulandığı evrenlerin örneklemleri arasındaki puan farkları görünen değer olarak alınmamalıdır. Araştırmacının farkları diğer deneysel yöntemlerle kanıtlama sorumluluğu vardır.

R.3: Karşıt kültürlerdeki karşılaştırmalar yalnızca rapor edilen puanlardaki değişmezlik düzeyinde yapılmalıdır.

R.4: Testi geliştiren evrende performansı etkileyecek sosyo-kültürel ve çevresel etkenlerin bilgisini elde etme yolu hakkında özel bilgi sağlamalıdır

ve sonuçların yorumlanmasında bu etkilerin hesaplanma işlemlerini önermelidir.

Bu rehber çeviri ve uyarlama alanında çalışan pek çok psikolog için yol göstericidir ve gelecek yıllarda artacak olan çeviri ve uyarlama ölçekleri için bunların genelleşmesi ve yaygınlaşması beklenmektedir. Bu rehberde, çevrilen ya da uyarlanan testlerin ya da araçların geçerliğini desteklemek için deneysel olduğu kadar yargısal derlemelerin önemi üzerinde durulmaktadır.

Ölçek uyarlamasında bu belirtilen rehber doğrultusunda uygun işlemler yapılmalıdır.

Ölçeğin psikometrik özellikleri (güvenirlik, geçerlik vb.) incelendikten sonra uyarlanan ölçeğin normları saptanır ve diğer dillerdeki ölçek normlarıyla karşılaştırılır. Öner (1987) bu aşamada şu sorulara yanıt arandığını bildirmektedir (Akt:Gözüm ve Aksayan, 2001):

- Uyarlanan ölçeğin puan ortalamaları, standart sapmaları ve kesme noktası gibi özellikleri asıl ölçeğin normlarına benziyor mu?
- Uyarlanan ölçeğin ölçme hatası asıl ölçeğin ölçme hatasına benziyor mu?
- Uyarlanan ve asıl ölçeğin faktör yapıları benziyor mu?
- Faktör-madde yükü her iki ölçekte de benzer değerlerde mi?

Bu özellikler zaten ölçeğin psikometrik özellikleriyle ortaya çıkacaktır. Kültürler arası karşılaştırma yapmak için her iki dilde de ölçeği geliştiren ve ölçeği uyarlayan araştırmacılar birlikte çalışmalılar. Her iki kültürde de benzer nitelikte olan bireylere uygulanan ölçekler karşılaştırılabilir. Sonra da kültürel özelliklerden kaynaklanan ayrılıklar tartışılır (Gözüm ve Aksayan, 2001).

SONUÇ

Pek çok psikolojik ve sosyolojik araştırmada standart ölçme aracı kullanılmaktadır. Standart ölçme aracıyla kast edilen, güvenirlik ve geçerlik açısından yeterli sayılabilecek objektif ölçme araçlarıdır. Standart ölçme araçlarının kullanıldığı araştırmalarda hem yapılan yorumlar daha objektif verilere dayanmaktadır hem de daha güvenilir sonuçlar ortaya konulmaktadır.

Bu tür araştırmalarda kullanılacak ölçek araştırmacı tarafından o araştırmaya özel olarak geliştirilebilir, daha önce kendi kültüründe ve dilinde geliştirilmiş olan ve psikometrik özellikleri yeterli olan bir ölçeği kullanabilir ya da bir başka kültürde geliştirilmiş olan güvenilir ve geçerli bir ölçeği kullanabilir. Bu konuda tercih araştırmacıya aittir. Kendi kültüründe

geliştirilmiş geçerli ve güvenilir bir aracı kullanmak bu tercihler içinde en az zaman ve emek gerektiren seçenektir. Ölçeği araştırmacının geliştirmesi veya başka bir ölçeği uyarlaması oldukça zahmetli ve zaman alıcı bir süreçtir.

Bu çalışmada psikolojik ölçme aracı uyarlama ve uyarlamanın aşamaları üzerinde durulmuştur. Uyarlamanın aşamaları doğru olarak izlenmediği takdirde yapılacak olan uyarlamanın ölçeğin asıl kültürdeki geliştiricisi tarafından reddedilmesi de söz konusu olabilmektedir.

Uyarlama konusunda yanlış inançlardan birisi uyarlamanın geliştirmeye göre daha kolay ve daha az zaman alıcı olduğudur. Halbuki, daha önce de değinildiği gibi uyarlama da en az geliştirme kadar zor ve zaman alıcıdır. Çoğu zaman bir ölçme aracını yeniden geliştirmek uyarlama yapmaktan kolay aşamalarla ilerlemektedir. Ancak bu, uyarlamanın olumsuz bir yönü olarak düşünülmemelidir. Gerektiğinde uyarlama önem kazanmakta, hatta kaçınılmaz bir hal almaktadır. Uyarlama özellikle kültürler arası psikolojik özelliklerin karşılaştırılması söz konusu olduğunda sık sık kullanılmaktadır.

KAYNAKÇA

Cronbach, L. J. (1990). *Essentials of psychological testing (5. Edition).* New York: Harper&Row Publishers.

Geske, A. & Andris, K. (2002). Evaluation of Latvia's science education in the IEA TIMSS and OECD PISA framework. *Journal of Baltic Science Education*, 2, 59-66.

Gözüm S. ve Aksayan S. (2003). Kültürlerarası ölçek uyarlama rehberi II. *Hemşirelikte Araştırma-Geliştirme Dergisi*, 5(1), 3-14.

Hambleton, R.K. & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology*, 1(1), 1-30.

Hambleton, R. K. & De Jong, J.H.A.L. (2003). Advances in translating and adapting educational and psychological tests. *Language Testing*, 20(2), 127-134.

INTESTCOM (2004). "About the international test commission." URL: www.intestcom.org.

Oakland, T. Poortinga, Y.H. Schlegel, J. & Hambleton, Y.K. (2001). International test commission: its history, current status and future directions**.** *International Journal of Testing*, 1(1), 3-32

OECD(2004). **"**OECD programme for international student assessment." URL:http://www.pisa.oecd.org/pisa/summary.htm.

PISA.OECD(2006). URL: http://www.pisa.oecd.org/pages/0,2987,en_ 32252351_32235731_1_1_1_1_1,00.html.

Sireci S.G. ve Berberoğlu G. (2000). Using bilingual respondents to evaluate translated-adapted items. *Applied Measurement in Education*, 13(3), 229-248.