

# Olumlu ve Olumsuz Duygulanımın Tekrarlanan Ölçümlerde Kişi-İçi Güvenirliği

Ali Can Gök

Orta Doğu Teknik Üniversitesi

Emre Selçuk

Orta Doğu Teknik Üniversitesi

Tülin Gençöz

Orta Doğu Teknik Üniversitesi

## Özet

Olumlu ile olumsuz duygulanım, boylamsal çalışmalarda ve danışan takibinde sıklıkla kullanılmasına rağmen, kullanılan ölçüm araçlarının kişi-İçi değişimi ne kadar güvenilir ölçtükleri henüz test edilmemiştir. Geleneksel test-tekrar test katsayısı temel varsayımları nedeniyle kişi-İçi değişimi yansıtmamaktadır, bu nedenle kullanılan ölçüm araçlarının hem kişi-İçi güvenirliginin hem de tekrarlanan ölçümlerde kişilerarası güvenirliginin test edilmesine ihtiyaç duyulmaktadır. Bu çalışma söz konusu iki soruyu Pozitif ve Negatif Duygu Ölçeği (PNDÖ) özelinde inceleyen ilk çalışmadır. Sırasıyla 4 tekrarlanan ölçüm ve 13 tekrarlanan ölçüm alınmış iki ayrı veri seti kullanılmıştır. İlk çalışma için 199 katılımcı 3 hafta ara ile PNDÖ'yü dört defa, ikinci çalışma için ise 41 katılımcı haftada bir kere olmak üzere PNDÖ'yü toplam 13 kere doldurmuştur. Cranford ve arkadaşlarının (2006) öne sürdüğü yöntem ile PDNÖ'nün kişi-İçi ve kişilerarası değişim güvenirlilik katsayıları hesaplanmıştır. İlk çalışma için kişi-İçi değişim güvenirlilik katsayısı Pozitif Duygu için .80, Negatif Duygu içinse .79 olarak bulunmuştur. Ayrıca ikinci çalışmada kişi-İçi değişim güvenirlilik katsayısı Pozitif Duygu için .86, Negatif Duygu içinse .87 olarak bulunmuştur. Ayrıca, her iki çalışma için kişilerarası güvenirlilik göstergeleri hesaplanmıştır. PDNÖ'nün her iki çalışmada da kişi-İçi ve kişilerarası değişimi uygun düzeyde ölçebildiği bulunmuştur. Boylamsal çalışmalarda PDNÖ'nün kişi-İçi ve kişilerarası değişim katsayılarına da yer verilmesi kuvvetle önerilmektedir.

**Anahtar kelimeler:** Duygu, duygulanım, pozitif ve negatif duygu ölçeği, kişi-İçi değişim, güvenirlilik, test-tekrar test katsayısı

## Abstract

Repeated positive and negative affect measurements are widely administered to adults in Turkey in both empirical and clinical work. However, within-person reliability of these measures has never been tested. Test-retest coefficient, a traditional index of reliability, does not evaluate within-person reliability and therefore, an investigation regarding the assessment of both within and between person reliability for repeated measures of positive and negative affect is crucial. The present paper examines the within-person reliability of the widely used Positive and Negative Affect Schedule (PANAS) in Turkish context. Two data sets were included. The first study included 196 participants, with three weeks' time interval at four time points. The second study included 41 participants, with one week time interval at thirteen time points. Within person reliability was calculated via the method recommended by Cranford et al. (2006). In the first study within-person reliability for Positive Affect was found as .80, and for Negative Affect as .79. In the second study, within-person reliability for Positive affect was found as .86, and for Negative Affect as .87. Furthermore, between-person reliability indices were calculated for each study. The PANAS had satisfactory within and between-person reliability coefficients in both studies.

**Keywords:** Affect, emotions, positive and negative affect scale, within-person change, test-retest coefficient

**Yazışma Adresi:** Dr. Ali Can Gök, Orta Doğu Teknik Üniversitesi, Fen-Edebiyat Fakültesi, Psikoloji Bölümü, Üniversiteler Mah. Dumlupınar Blv. No:1, Çankaya / Ankara

**E-posta:** alicangok@gmail.com

**Gönderim Tarihi:** 20.03.2017

**Kabul Tarihi:** 29.10.2017

Duygular, psikolojik işlevselliği değerlendirilmede özellikle tekrar edilen ölçümler kullanılarak gerek bilimsel araştırmalarda gerekse klinik uygulamalarda kullanılan önemli bir olgudur (Merz ve Roesch, 2011; Röcke, Li ve Smith, 2009). Popüler olarak “günlük yöntemleri” olarak adlandırılan deneyim örnekleme çalışmalarında, araştırmacı katılımcılardan belirli zaman aralıklarında (örn., günlük, haftalık, aylık) ölçüm olarak olumlu ve olumsuz duygulanımın nasıl değiştiğini veya dalgalandığını inceler. Örneğin, bu yöntem aracılığıyla elde edilen bulgulara göre, gün içinde yaşanan stresli olayların duygulanımı olumsuz yönde etkilediği (Bolger, DeLongis, Kessler ve Schilling, 1989), yaşanan olumlu olayların ise olumlu yönde etkilediği bulunmuştur (Zautra, Affleck, Tennen, Reich ve Davis, 2005). Bunun yanı sıra, deneyim örnekleme çalışmalarıyla kadınların doğum kontrol hapı kullanımının duygulanım üzerinde (Oinonen ve Mazmanian, 2001) ve ergenlerde duygulanımın riskli cinsel davranış üzerinde (Shrier, Shih ve Beardslee, 2005) etkisi bulunmuştur. Yukarıda örnek verilen çalışmalarda duygulanımın gidişatının ve iniş çıkışlarının da araştırılan konular üzerindeki etkisi duygu ölçümlerinin tekrar tekrar alınması ile ortaya çıkarılmıştır.

Deneyim örnekleme çalışması kapsamına girmeyen fakat tekrar eden ölçüm alınan boylamsal çalışmalarda da duygular sıkça araştırılmaktadır. Duygu ölçümünün iki farklı zaman noktasında yapıldığı çalışmalarda; ergenlerde kişilik özelliklerinin duygulanım ve buna bağlı davranışlar ile (Ciarrochi ve Heaven, 2007), uzun süren işsizlik durumunun duygulanım ile (Ferreira ve ark., 2015), metabolik sağlığın duygulanım ile (Jeffery, Hyland, Hosking ve Wilkin, 2014), olumlu duygulanımın olumsuz duygulanım ile (Moneta, Vulpe ve Rogaten, 2012), duygulanımın geç evre kanser ile (Rabkin, McElhiney, Moran, Acree ve Folkman, 2009) ve erken evre meme kanseri hastalarında yoganın duygulanım ile (Vadira ve ark., 2009) ilişkileri bulunmuştur. Duygu ölçümünün üç defa veya daha fazla tekrarlandığı çalışmalarda; duygulanım ile materyalizm ve disiplin kültürünün (Jiang, Song, Ke, Wang ve Liu, 2009), duygulanım ile üniversiteye giriş sınavına hazırlanma sürecinin (Peluso, Savalli, Curi, Gorenstein ve Andrade, 2009), Alzheimer hastalarının eşlerinde uyku ile duygulanım (von Kanel ve ark., 2014) arasında ilişkiler bulunmuştur.

Duygu değerlendirmesinde araştırmacılar önce duyguların üzerinde tek tek durmuşlar; üzüntü, korku, kaygı, öfke ve mutluluk gibi duyguları ayrı ayrı ele almışlardır (bkz. Ekman 1984; Izard 1972, 1977). Daha sonra, duygu araştırmalarında duygular boyutlar üzerine yerleştirilmeye başlanmıştır (Watson ve Tellegen, 1985; Watson ve Vaitya, 2003). Duygular boyutsal olarak incelenirken ilk önce tek bir boyutun olduğu ve olumlu duyguların bir uçta, olumsuz duyguların da diğer uçta

olduğu, böylece bir duygu etkinken bu duygu ile uyumsuz olan diğer duyguların aktif olmadığı bir yapı olduğu savunulmuştur (Barrett ve Russell, 1998; Russell ve Barrett, 1999; Watson ve Vaitya, 2003). Bu bakış açısıyla değerlendirildiğinde, duyguların olumlu olumsuz uçlardan oluşan tek boyutun üzerinde yer aldığı ve birinin varlığının diğerinin yokluğunu işaret ettiği düşünülmektedir (Merz ve Roesch, 2011).

İlerleyen çalışmalar, olumlu ve olumsuz duyguların aynı boyutun iki ucu olmaktan ziyade, birbirinden bağımsız iki boyut (ortagonal) üzerinde değerlendirilmesi gerektiğini savunmuştur (Merz ve Roesch, 2011; Watson, Clark ve Carey, 1988). Bu ortogonal yapı çerçevesinde aynı anda hem olumlu hem de olumsuz duyguların aktif olabileceği öngörülmekte, böylece birinin varlığı diğerinin yokluğu ile olabileceği varsayımı çürütülmektedir (Watson ve ark., 1988).

Olumlu ve olumsuz duygulanımı ölçmek için Pozitif ve Negatif Duygu Ölçeği (PNDÖ) (Watson ve ark., 1988) en yaygın olarak kullanılan ölçüm aracıdır (Schmukle, Egloff ve Burns, 2002; Watson ve Vaitya, 2003). Kısaca özetlemek gerekirse PNDÖ olumlu duygulanımı, bireyin hissettiği hevesli, aktif ve uyanık olma durumlarını kapsayacak şekilde pozitif duygu alt ölçeği ile ölçmektedir (Watson ve ark., 1988). Pozitif duygu alt ölçeğinden yüksek puanlara sahip olmak yüksek enerji, zevk alarak ve yoğunlaşarak iş yapabilme durumunu yansıtmaktadır. Öte yandan pozitif duygu alt ölçeğinden düşük puanlara sahip olmak ise durgunluk ve üzüntü olarak özetlenebilir. Negatif duygu alt ölçeği ise öznel stresin yanı sıra öfke, korku, suçluluk gibi olumsuz duyguları içermektedir. Negatif duygu alt ölçeğinden düşük puan alınması ise sakinlik ve huzur hissetme durumunu yansıtmaktadır (Watson ve ark., 1988; Gençöz 2000).

PNDÖ birçok çalışmada kullanılmış; algılanan stres ve sağlık sorunları (Watson ve Pennebaker, 1989), kaygı ve depresif yakınmalar (Gençöz 2000; Watson ve ark., 1988), ayrımcılığa maruz kalma (Brondolo ve ark., 2000), sosyal aktiviteler (Watson, Clark, McIntyre ve Hamaker, 1992), iş arama davranışı (Burger ve Caldwell, 2000), psikolojik dayanıklılık (Quale ve Schanke, 2010; Tugade ve Fredrickson, 2004), sigara tüketimi (Becona, Vazquez, Fuentes ve del Carmen Lorenzo, 1998) gibi çeşitli değişkenlerle ilişkili bulunmuştur. Bunun yanı sıra bu ölçek birçok dile de (örn., Türkçe, Almanca, Sırpça, İspanyolca) uyarlanmıştır (Gençöz 2000; Krohne, Egloff, Kohlman ve Tausch, 1996; Mihic, Novovic, Colovic ve Smederevac, 2014; Sandin ve ark., 1999).

PNDÖ, duyguların bir yandan durumluk, diğer yandan da süreklilik boyutunu ölçmek için kullanılmaktadır (Watson ve Clark, 1984; Watson ve ark., 1988). Süreklilik duygu, ortamlara göre nispeten daha az değişen duygusal yatkinlik anlamına gelmekteyken; durumluk

duygu ise günlük olaylar ve bağlamsal özellikler neticesinde dalgalanabilen daha anlık duygusal deneyimleri yansıtmaktadır (Brondolo ve ark., 2008). Durumluk yada süreklilik duygulanımı ölçmek amacıyla yapılan ölçümlerde yönergeler değiştirilerek değerlendirilmek istenen zaman dilimi belirtilebilmektedir (Watson ve ark., 1988). Örneğin süreklilik duyguyu ölçmek amacıyla yönerge “genel olarak düşündüğünüzde kendinizin nasıl hissettiğini değerlendiriniz” şeklindeyken; günlük duyguyu ölçmek amacıyla “bugün nasıl hissettiğinizi değerlendiriniz” olarak değiştirilmektedir. Böylece katılımcının genel olarak ya da belirlenen kısa süreye özel olarak duygularını değerlendirmeleri istenmektedir (Merz ve Roesch, 2011). PNDÖ’yü uygularken verilen yönergeye göre “şu an, bugün, son günlerde, geçen hafta, geçen haftalarda, geçen ay, geçen yıl ve genel olarak” şeklinde sekiz ayrı zaman dilimine yönelik ölçüm almak mümkündür (Watson ve ark., 1988). Ölçüm almak istediğimiz zaman dilimine göre yönergeyi değiştirmek duyguları ölçerken durumluk ya da süreklilik duygular arasındaki farka duyarlı gözükse de, kişinin geçmişindeki bir duygulanıma yönelik tek bir ölçüm alındığı durumlarda bu değerlendirme geriye dönük hatırlamada oluşacak yanlışlıklara ve hatalara maruz kalabilmektedir (Merz ve Roesch, 2011). Bu nedenle, bu tip değerlendirmelerde duygulanımdaki varyansın ne kadarının durumluk etkiden, ne kadarının süreklilikten kaynaklandığını ayırtlamak oldukça zordur. Bu yanlışlıktan kurtulmak için geçmişe yönelik değerlendirmeler almak yerine, durumluk duyguları farklı zamanlarda tekrarlanan ölçümlerle almak, durumluk ve süreklilik duyguyu ayırtlamak için en ideal yöntem olarak görülmektedir (Eid ve Diener, 1999; Merz ve Roesch, 2011).

Her ne kadar pozitif ile negatif duygu alt ölçekleri ortogonal yapılarla sahip olsa da, yapılan çalışmalar bu iki boyutun arasında orta düzeyde ters yönde bir ilişki olduğunu ortaya koymaktadır (Crawford ve Henry, 2004; Lonigan, Phillips ve Hooe, 2003; Tellegen, Watson ve Clark, 1999). Ayrıca, pozitif duygu ile negatif duygu alt ölçeklerinin arasındaki ilişkinin büyüklüğü, hedef alınan zaman dilimine göre de değişim gösterebilmektedir (Schmukle ve ark., 2002). Pozitif ve negatif duygu alt ölçekleri arasındaki ilişki daha kısa zaman dilimleri hedef alındığında, uzun zaman dilimlerine kıyasla daha kuvvetli çıkmaktadır (Diener ve Emmons, 1984). Bu nedenle, durumluk olarak bakıldığında pozitif ve negatif duygu alt ölçekleri arasındaki ilişki daha kuvvetli gözükmektedir. Önerilen bu tekrar ölçümlerle değerlendirme yönteminde de süreklilik duygunun durumluk duygu üzerine etkisi olabilir, yani duygulanımdaki genel yatkınlık bir noktaya kadar durumluk duyguyu da yordamaktadır (Schmukle ve ark., 2002), ancak bu etkinin daha zayıf olması beklenmektedir (Brondolo ve ark., 2008).

Durumluk ve süreklilik duyguyu ayırttırmanın yanı sıra, tekrarlanan ölçümler aynı zamanda bireyin kendi içindeki duygusal değişimleri anlamak için de kullanılır (Cranford ve ark., 2006; Eid ve Diener, 1999). Bu tür araştırma desenlerinde katılımcılardan günlük, haftalık veya aylık olarak düzenli ölçümler alınır, bu yöntem geriye dönük hatırlamadan kaynaklı hatalara ve yanlışlıklara daha az mahal verdiği için, kişinin değişimi yaşadığı ana dair bilgiyi daha hatasız bir şekilde alma imkânı sağlar (Cranford ve ark., 2006). Katılımcılardan farklı zamanlarda peş peşe alınan ölçümler, söz konusu olguyu kendi doğal ortamında ve bağlamında aradan çok zaman geçmeden incelemeye izin verdiği için araştırmacıya geleneksel yöntemlerin sağlayamadığı kadar ayrıntılı bir tablo sunma avantajı sağlamaktadır (Bolger, Davis ve Rafaeli, 2003; Reis, 1994). Ancak bu zengin ve ayrıntılı tablonun güvenilirliği, tekrarlanan ölçümlerde kullanılan aracın kişi-İçi değişimi ne kadar hassas ölçtüğüne bağlıdır.

### Kişi-İçi Güvenirlik

Duyguların tekrarlı bir şekilde ölçüldüğü bu desenlerde dikkat edilmesi gereken kritik nokta, ölçülen farklılıkların iki kaynağı olmasındır; kişilerarası farklar ve kişi-İçi farklar. Duygu durumunun tek seferde ölçüldüğü geleneksel desenler katılımcıların birbirinden farkını ortaya koymaktadır. Nitekim duygu durumunu ölçmek için en sık kullanılan araçlardan biri olan PNDÖ için şimdiye kadar yapılmış güvenilirlik çalışmaları da PNDÖ’nün kişilerarası farkları ne kadar güvenilirlikte yakaladığını incelemiştir (Gençöz, 2000; Watson ve ark., 1998). Ancak duygu durumu tekrarlı bir şekilde ölçüldüğü zaman elde edilen varyansın tek kaynağı kişilerarası farklar değil, aynı zamanda kişinin kendi içindeki değişimlerdir (Bolger ve ark., 2003; Cranford ve ark., 2006; Hu ve ark., 2016). Fiske ve Rice’a (1955) göre kişi-İçi varyansı doğru bir şekilde analiz edebilmek için o kişiden elde edilen test ya da test maddelerindeki puanlardaki iniş çıkışların sistematik olmayan ölçüm hatasından ayıklanması gerekmektedir. Bu nedenle, bir kişiden farklı zamanlarda alınan ölçümlerdeki farklılıkların ne kadarının sistematik olmayan ölçüm hatası olduğunu ve ne kadarının da gerçekten ölçüm alınan kişide yaşanan değişiklikten kaynaklandığını bilmek kişi-İçi değişim süreçlerini çalışırken göz önünde bulundurulması gereken en kritik noktadır (Eid ve Diener, 1999). Dolayısıyla, kullanılan ölçüm aracı için kişinin kendi içindeki değişimi ne kadar tutarlı olarak ölçebildiğinin bilinmesi gerekmektedir (Fiske ve Rice, 1955).

Genel olarak, ölçek geliştirilirken yahut başka bir lisana/kültüre uyarlanırken test-tekrar test katsayısı rapor edilir. Fakat bu katsayı, söz konusu ölçeğin iki farklı zaman noktasında aynı katılımcılara uygulanması neticesinde elde edilen test puanlarının birbirleriyle korelasyonundan ibarettir (Guttman 1945; Hu ve ark., 2016;

**Tablo 1.** Birinci Çalışmanın Tanımlayıcı İstatistikleri

Ölçüm	n	Katılım Oranı	Cinsiyet Dağılımı			Pozitif Duygu		Negatif Duygu	
			Kadın	Erkek	Belirtmemiş	Ort.	S.	Ort.	S.
1	196	%98	%55.00	%39.39	%5.61	29.13	7.37	21.61	7.45
2	167	%83	%56.29	%41.32	%2.39	28.30	7.05	20.56	7.08
3	131	%66	%61.83	%38.17	-	28.25	7.98	20.33	6.27
4	94	%47	%60.64	%39.36	-	26.75	8.60	21.21	7.37

Shrout ve Lane, 2012; Traub 1994). Test-tekrar test katsayısı, bir kişinin ölçekten aldığı puanı gerçek puan ve sistematik olmayan hata payının toplamı olarak ele alır. Yapılan iki ölçümde kişinin gerçek puanının değişmeyeceği varsayımı olduğu için eğer söz konusu kişi iki ölçümde farklı puan aldıysa bu aradaki fark sistematik olmayan hata payı olarak değerlendirilir. Bu noktada da iki ölçüm arasındaki korelasyonun ölçüm alınan kişilerin gerçek puanlarının tutarlılığını yansıttığı düşünülmektedir (Shavelson, Webb ve Rowley, 1989). Bu katsayı, deneyim örnekleme çalışırken veya değişimi ölçmeyi hedefleyen boylamsal çalışma yapılırken problemleri olabilir, çünkü ölçüm alınan farklı zaman noktalarındaki sistematik olmayan ölçüm hatası ile gerçek kişi-içi değişim ayrıştırılmamıştır (Hu ve ark., 2016; Eid ve Diener, 1999). Yani bir kişinin iki ölçümden farklı puan aldığı bir durumda test-tekrar test katsayısı bu farkı sistematik olmayan hata olarak değerlendirmektedir. Örneğin, bir ölçeğin test-tekrar test katsayısı düşük çıktığında bunun iki sebebi olabilir. Birincisi, farklı zamanlarda ölçekten alınan puanlar gerçekten kişinin kendi içindeki değişimi yansıtmaktadır, yani kişi ölçeğin ölçtüğü olguda bir değişim yaşamıştır ve düşük test-tekrar test katsayısı bu değişimi yansıtmaktadır (Hu ve ark., 2016). İkincisi ise, kişide anlamlı bir değişim yoktur, fakat ölçüm aracı güvenilir bir şekilde söz konusu olguyu ölçmemiştir (Fiske ve Rice, 1955). Bu durumda, veriyi yuvalanmış bir şekilde ele almadan, yani aynı kişiden farklı zamanlarda elde edilen test puanlarını birlikte değerlendirmeden, iki farklı zaman diliminde alınan ölçümlerdeki farklılıkların tam olarak nereden kaynaklandığını bilmek maalesef mümkün değildir (Raudenbush ve Bryk, 2002).

Bu sorunu aşmak için Cranford ve arkadaşları (2006) Genellebilirlik Kuramı çerçevesinde (Cronbach, Gleser, Nanda ve Rajaratnam, 1972) günlük araştırma desenlerinde kullanılacak ölçümlerin psikometrik özelliklerine dair fikirler sunmuşlardır. Genellebilirlik Kuramı, bir ölçekten elde edilen puanlarda farklı kaynaklardan gelen varyansların olabileceğini kabul eder (Shavelson ve ark., 1989). Bu durum, tek faktörlü ANOVA ve çok faktörlü ANOVA'nın arasındaki fark ile benzeşmektedir. Tek faktörlü ANOVA'da varyans, sistemat-

tik/gerçek varyans ve hata/rastgele varyans olarak ikiye ayrılmaktadır. Fakat çok faktörlü ANOVA'da varyans birden fazla faktörden, bunların etkileşiminden ve hata/rastgele varyansın oluşabilirden (Shavelson ve ark., 1989). Genellebilirlik Kuramı da çok faktörlü ANOVA ile varyansın birçok kaynaktan gelebileceğini kabul etmesi açısından benzeşmektedir (Shavelson ve ark., 1989). Test-tekrar test katsayısı hesaplanırken kullanılan kısıtlayıcı varsayımın aksine, Cranford ve arkadaşlarının (2006) öne sürdüğü yöntemde kişinin farklı zamanlarda alacağı puanın aynı olacağına dair bir varsayım yoktur. Bu da kişi-içi değişimin ölçülebileceğini ve sistematik olmayan hata payı olarak değerlendirilmemesi gerektiğini öngörmektedir (Shrout ve Lane, 2012).

Genellebilirlik Kuramı yaklaşımı pek çok açıdan araştırmacıya fayda sağlamaktadır. Bunlardan ilki geleneksel güvenilirlik yaklaşımlarını da kapsamı ve bunun ötesine geçip kişi-içi değişimin araştırıldığı deneyim örnekleme yöntemlerinde yarar sağlayabilecek ve anlamlı olacak tekrar eden ölçümlerdeki varyansın hangi kaynaklardan geldiğini ayırt edebilmesidir. Bu varyans bileşenleri, tekrar eden ölçümlere dayalı yöntemlerde kullanılabilir çeşitli güvenilirlik türlerini hesaplamada kullanılabilir (Cranford ve ark., 2006). Bu çalışmada kişi-içi güvenilirliğe ve üç ayrı kişilerarası güvenilirlik kat-sayısına yer verilecektir. Bunun yanı sıra, bahsi geçen katsayıları hesaplama yöntemi de tanıtılacaktır.

PNDÖ için kişi-içi güvenilirlik çalışmasının yapılması hem yukarıda belirtildiği gibi sıklıkla kullanıldığı boylamsal çalışmalarda güvenilir bir ölçüm aracı olup olmadığının test edilmesini hem de durumluk süreklilik duyguların daha sağlıklı bir şekilde ayrıştırılmasını sağlayacaktır.

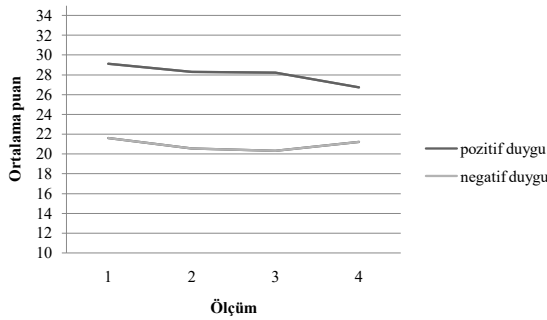
## Yöntem

### Genel Bakış

Bu çalışmada iki ayrı veri kaynağından gelen çalışma rapor edilmektedir. İlk PNDÖ'nün Türkiye uyarılmasında (Gençöz, 2000) toplanılan 4 kez ölçüm alınmış olan veri seti (Çalışma 1), diğeri ise ilk yazarın doktora tezi amacıyla yapılan bir çalışma kapsamında elde edilen 13 haftalık veri setidir (Çalışma 2).

## Örneklem

**Çalışma 1.** Çalışmaya 78'i (%39.2) erkek, 108'i (%54.3) kadın, 13'ü (%6.5) cinsiyet belirtmemiş toplam 199 üniversite öğrencisi katılmıştır (her bir ölçüme katılım oranı ve tanımlayıcı istatistikler için bkz. Tablo 1; alınan ortalama puanlar için bkz. Şekil 1). Katılımcıların yaşları 18 ile 34 arasında değişmektedir ( $Ort. = 20.90, S = 1.88$ ). Her katılımcının en az 2 ölçümü vardır. Ölçümlerin arasında üçer hafta vardır.



**Şekil 1.** Çalışma 1'de Alınan Ortalama Puanlar

**Çalışma 2.** Çalışmaya Ankara'da klinik psikoloji lisansüstü programlarında eğitim görüp psikoterapi süpervizyonu alan 46 terapist katılmış fakat minimum 2 hafta ölçüm alınması gerektiğinden çalışmaya 41 katılımcı ile devam edilebilmiştir. Katılımcıların 35'i (%85.4) kadın, 6'sı (%14.6) erkektir (her bir ölçüme ka-

tılım oranı ve tanımlayıcı istatistikler için bkz. Tablo 2; alınan ortalama puanlar için bkz. Şekil 2.). Katılımcıların yaşları 23 ile 29 arasında değişmektedir ( $Ort. = 25.56, S = 1.45$ ). Katılımcılardan birbirini takip eden haftalarda 13 kez ölçüm alınmıştır. Ölçümlerin arasında birer hafta bulunmaktadır.

## Veri Toplama Araçları

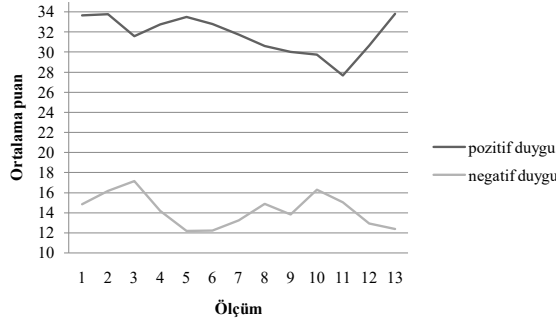
**Pozitif ve Negatif Duygu Ölçeği (PNDÖ).** Pozitif ve Negatif Duygu Ölçeği, Watson ve arkadaşları tarafından (1988) olumlu ve olumsuz duygulanımı değerlendirmek amacıyla geliştirilmiştir. Ölçeğin olumlu duygulanımı ölçen pozitif duygu ve olumsuz duygulanımı ölçen negatif duygu olmak üzere iki adet alt ölçeği bulunmaktadır. Ölçekte her bir alt ölçek için onar madde olmak üzere toplam 20 madde bulunmaktadır. Ölçek 1 (çok az – hiç) ve 5 (çok fazla) olmak üzere beşli Likert tipi değerlendirmeler üzerinden puanlanmaktadır. Böylece bir kişinin pozitif veya negatif duygu alt ölçeklerinden alacağı puan 10 ile 50 arasında değişmektedir. Ölçeğin orijinal formunun iç tutarlık katsayıları pozitif duygu için .88, negatif duygu için ise .85'tir. Test-tekrar test katsayısı ise .47'dir.

Ölçeğin Türkçe uyarlaması Gençöz (2000) tarafından gerçekleştirilmiştir. İç tutarlık katsayısı pozitif duygu için .86, negatif duygu için .83 bulmuştur. Test-tekrar test katsayısı ise pozitif duygu için .54, negatif duygu için .40'tır. Bunun yanı sıra pozitif duygunun depresyonla ( $r = -.48, p < .01$ ) ve kaygı ile ( $r = -.21, p < .05$ ) ters yönde ilişkisi; negatif duygunun ise depresyon ( $r = .51, p$

**Tablo 2.** İkinci Çalışmanın Tanımlayıcı İstatistikleri

Ölçüm	n	Katılım Oranı	Cinsiyet Dağılımı		Pozitif Duygu		Negatif Duygu	
			Kadın	Erkek	Ort.	S.	Ort.	S.
1	40	%98	%85	%15	33.67	6.33	14.86	4.07
2	32	%78	%91	%9	33.76	6.27	16.15	5.64
3	38	%93	%87	%13	31.58	6.47	17.16	7.97
4	33	%81	%91	%9	32.76	5.82	14.16	4.94
5	30	%73	%94	%6	33.50	6.55	12.17	2.82
6	33	%80	%91	%9	32.79	5.89	12.19	2.43
7	27	%66	%89	%11	31.74	7.26	13.22	3.49
8	26	%64	%88	%12	30.61	8.49	14.89	6.04
9	31	%76	%90	%10	30.03	7.25	13.81	4.43
10	29	%70	%90	%10	29.72	7.68	16.31	7.96
11	19	%46	%90	%10	27.68	6.97	15.05	4.35
12	20	%48	%90	%10	30.65	5.84	12.90	3.37
13	16	%39	%100	-	33.81	6.53	12.38	3.22

< .01) ve kaygı ile ( $r = .47, p < .01$ ) pozitif yönde ilişkisi gösterilmiştir (Gençöz, 2000). Bu psikometrik özellikler ölçeğin Türkçe formunun güvenilirlik ve geçerliğini desteklemektedir.



Şekil 2. Çalışma 2’de Alınan Ortalama Puanlar

### İşlem

Yukarıda bahsedildiği gibi Genelilebilirlik Kuramı (Shavelson ve ark., 1989) elde edilen varyansın farklı kaynaklardan gelebileceğini kabul eder. Genelilebilirlik Kuramı çerçevesinde bir katılımcı tarafından bir maddeye bir zaman noktasında verilen puan varyans bileşenlerine ayrılabilir (Cranford ve ark., 2006). Varyans bileşenleri katılımcı, zaman noktası ve madde olmak üzere üç faktör ve etkileşimleri üzerinden ayrıştırılabilir. Örneğin,  $j$  kişisi için  $i$  maddesine  $k$  zaman noktasında verdiği puan ( $P$ ):

$$P_{ijk} = \mu + K_j + M_i + Z_k + (KM)_{ij} + (KZ)_{jk} + (MZ)_{ik} + (KMZ)_{ijk} + e_{ijk}$$

şeklinde modellenenmektedir. Tüm ölçeğin tüm zamanlardaki tüm maddelerinin ortalaması  $\mu$  ile ifade edilmektedir.  $K_j$ ,  $j$  kişinin tüm maddelerde ve zaman noktalarındaki etkisini;  $M_i$ ,  $i$  maddesinin tüm kişi ve zaman noktalarındaki etkisini;  $Z_k$ , ise  $k$  zaman noktasının tüm kişi ve maddeler üzerindeki etkisini göstermektedir. Bunların yanı sıra,  $(KM)_{ij}$ ,  $i$  maddesi ve  $j$  kişinin tüm zaman noktalarındaki etkisini;  $(KZ)_{jk}$ ,  $j$  kişinin ve  $k$  zaman noktasının tüm maddeler üzerindeki etkisini;  $(MZ)_{ik}$ , ise  $i$  maddesinin ve  $k$  zaman noktasının tüm kişiler üzerine etkisini ifade etmektedir.  $(KMZ)_{ijk}$  ve  $e_{ijk}$  sembolleri ise  $i$  maddesinin  $j$  kişisi için  $k$  zaman noktasındaki etkilerini göstermektedirler.  $(KMZ)_{ijk}$  sistemli bir etkiyi ifade ederken,  $e_{ijk}$  sistemli olmayan bir etkiyi ifade etmektedir (Cranford ve ark., 2006). Yukarıda bahsedilen tüm bu etkileri, tüm ölçeğin tüm zamanlardaki tüm maddelerinin ortalaması  $\mu$ 'ye ekledikten sonra  $j$  kişinin  $i$  maddesine  $k$  zaman noktasında verdiği puana ulaşabiliriz.

Yukarıda bahsedilen bu etkilerin varyansları SPSS'in VARCOMP prosedürü ile hesaplanabilir (ilgi-

li SPSS kodu için Ek-1'e bakınız). Varyans bileşenleri hesaplandıktan sonra zaman içindeki kişi-içi sistematik değişimin güvenilirlik katsayısı için:

$$G_d = \sigma^2_{\text{Kişi*Zaman}} / [\sigma^2_{\text{Kişi*Zaman}} + (\sigma^2_{\text{Hata}} / m)]$$

formülü uygulanmaktadır. Kişilerin zaman içinde gösterdikleri varyans, kişilerin zaman içerisinde gösterdiği varyans ile hata bileşeninin madde sayısına bölümünün toplamına bölünmüştür (Cranford ve ark., 2006). “ $m$ ” madde sayısını ifade etmektedir.

Bunun yanı sıra, tekrar eden ölçümler alınan çalışmalarda kişilerarası güvenilirlik de tek ölçüm alınan çalışmalara nazaran farklılıklar göstermektedir (Cranford ve ark., 2006). Tek ölçüm alınan çalışmalarda verilen Cronbach alfa değeri bir zamandaki ölçüm için geçerlidir. Fakat birden fazla ölçüm alındığında kişilerarası güvenilirliğin birden fazla göstergesini hesaplamak mümkündür (Cranford ve ark., 2006). Aşağıda üç adet kişilerarası güvenilirlik göstergesi açıklanacaktır.

Bunlardan ilki birden fazla ölçüm alınan bir çalışmada, tek bir ölçümde beklenebilecek kişilerarası güvenilirlik katsayısıdır (Cranford ve ark., 2006). Bunun için formül:

$$G_1 = [\sigma^2_{\text{Kişi}} + (\sigma^2_{\text{Kişi*Madde}} / m)] / [\sigma^2_{\text{Kişi}} + (\sigma^2_{\text{Kişi*Madde}} / m) + (\sigma^2_{\text{Hata}} / m)]$$

şeklinde dir. Kişilerden gelen varyans ve kişilerin maddeler arasında gösterdiği varyansın madde sayısına bölümü ile toplamının; kişilerden gelen varyans, kişilerin maddeler arasında gösterdiği varyansın madde sayısına bölümü ve hata bileşeninin madde sayısına bölünmesinin toplamının bölümü ile elde edilmektedir. Bu formül birden fazla ölçümü olan çalışmanın herhangi tek bir ölçümündeki kişilerarası güvenilirlik katsayısını vermektedir (Cranford ve ark., 2006).

İkinci kişilerarası güvenilirlik katsayısında ise tekrar eden ölçümlere sahip bir çalışmanın tek ölçümünden ziyade tüm ölçümlerindeki kişilerarası güvenilirliği verilmektedir. Bu katsayı diğerlerine göre daha düşük çıkmaktadır çünkü ilk formülde A kişinin X zamanındaki ölçümü ile B kişinin X zamanındaki ölçümlerinin arasındaki güvenilirliğe bakılırken burada ölçüm alınan zaman noktaları rastgele bırakılmakta ve A kişinin X zamanındaki ölçümü ile B kişinin Y zamanındaki ölçümü arasındaki kişilerarası güvenilirlik katsayısına bakılmaktadır. Formül:

$$G_2 = [\sigma^2_{\text{Kişi}} + (\sigma^2_{\text{Kişi*Madde}} / m)] / [\sigma^2_{\text{Kişi}} + (\sigma^2_{\text{Kişi*Madde}} / m) + \sigma^2_{\text{Zaman}} + \sigma^2_{\text{Kişi*Zaman}} + (\sigma^2_{\text{Hata}} / m)]$$

şeklinde dir. Burada bir önceki formüle kıyasla payda kısmına zaman varyansı ve kişilerin zaman içerisinde gösterdikleri varyans eklenmiştir. Bu formül tekrar eden ölçümleri olan bir çalışmada herhangi bir kişiden herhangi bir zamanda alınmış ölçüm ile başka bir kişiden herhangi bir zamanda alınmış olan ölçümler arasındaki kişilerarası güvenilirlik katsayısını vermektedir (Cranford ve ark., 2006).

**Tablo 3.** Birinci Çalışmanın Varyans Bileşenleri Analizi Sonuçları

Varyans Bileşeni	Pozitif Duygu	%	Negatif Duygu	%
$\sigma^2_{\text{Kişi}}$	.269	19.75	.219	13.94
$\sigma^2_{\text{Madde}}$	.048	3.52	.368	23.42
$\sigma^2_{\text{Zaman}}$	.005	0.37	.002	0.13
$\sigma^2_{\text{Kişi*Madde}}$	.248	18.21	.219	13.94
$\sigma^2_{\text{Kişi*Zaman}}$	.230	16.89	.210	13.37
$\sigma^2_{\text{Madde*Zaman}}$	.004	0.29	.002	0.13
$\sigma^2_{\text{Hata}}$	.558	40.97	.551	35.07
<b>Toplam</b>	<b>1.362</b>	<b>100</b>	<b>1.571</b>	<b>100</b>

Üçüncü ve son kişilerarası güvenilirlik katsayısında ise kişilerin tekrar eden ölçümlerden aldığı ortalama puanlar arasındaki kişilerarası güvenilirlik katsayısı hesaplanmaktadır. Bu katsayı, A kişinin  $k$  adet ölçümden aldığı puanların ortalaması ile B kişinin  $k$  adet ölçümden aldığı puanların ortalamasının arasındaki farkı ölçeğin ne tutarlılıkta değerlendirdiğini göstermektedir. “ $k$ ” tekrar edilen ölçüm sayısı olmak üzere, formül:

$$G_3 = \frac{[\sigma^2_{\text{Kişi}} + (\sigma^2_{\text{Kişi*Madde}} / m)]}{[\sigma^2_{\text{Kişi}} + (\sigma^2_{\text{Kişi*Madde}} / m) + (\sigma^2_{\text{Hata}} / k * m)]}$$

şeklinde dir. Kişilerden gelen varyans ve kişilerin maddeler arasında gösterdiği varyansın madde sayısına bölümü ile toplamının; kişilerden gelen varyans, kişilerin maddeler arasında gösterdiği varyansın madde sayısına bölümü ve hata bileşeninin tekrar eden ölçüm sayısı ( $k$ ) ile madde sayısının çarpımına bölümünün toplamının bölümü ile elde edilmektedir. Bu formülde payda kısmına alınan ölçüm sayısı ile bir düzenleme yapıldığı için daha fazla ölçüm almış olmak, bu kişilerarası güvenilirlik göstergesini yükseltmektedir. Dolayısıyla daha çok ölçüm almış olmak, daha yüksek kişilerarası güvenilirlik sağlamaktadır. Bu durum da kişilerin sürekli karakteristiklerini daha güvenilir bir şekilde ölçülmesine imkan vermektedir.  $G_3$ , ölçüm sayısını da dikkate alan ve kişinin katıldığı ölçümlerin ortalaması temelinde hesaplanan bir gösterge olduğu için kritik ve önemli bir katsayıdır.

### Bulgular

İlk çalışmaya varyans bileşenleri analizi uygulandığında elde edilen sonuçlar Tablo 3'te gösterilmiştir. Değişim güvenilirliği ( $G_d$ ) için yukarıda belirtilen formül uygulandığında elde edilen sonuç pozitif duygu için .80 iken negatif duygu için .79 olarak bulunmuştur. İlk çalışmada aynı zaman noktası için kişilerarası güvenilirlik katsayısı ( $G_1$ ) pozitif duygu için .84 iken negatif duygu için .81 bulunmuştur. Tüm zaman noktaları için kişilerarası güvenilirlik ( $G_2$ ) pozitif duygu için .50, negatif duygu için .47 olarak bulunmuştur. Kişilerin tüm ölçüm noktalarındaki ortalama puanları üzerinden güvenilirlik katsayısı ise ( $G_3$ ) pozitif duygu için .95, negatif duygu için .95 bulunmuştur.

Buna ek olarak, ikinci çalışmaya varyans bileşenleri analizi uygulandığında elde edilen sonuçlar Tablo 4'te gösterilmiştir. Değişim güvenilirliği ( $G_d$ ) için yukarıda belirtilen formül uygulandığında elde edilen sonuç pozitif duygu için .86 iken negatif duygu için .87 olarak bulunmuştur. İkinci çalışmada aynı zaman noktası için kişilerarası güvenilirlik katsayısı ( $G_1$ ) pozitif duygu için .86 iken negatif duygu için .67 bulunmuştur. Tüm zaman noktaları için kişilerarası güvenilirlik ( $G_2$ ) pozitif duygu için .39, negatif duygu için .18 olarak bulunmuştur.  $G_2$  hesaplanırken zaman noktaları rastgele bırakıldığından

**Tablo 4.** İkinci Çalışmanın Varyans Bileşenleri Analizi Sonuçları

Varyans Bileşeni	Pozitif Duygu	%	Negatif Duygu	%
$\sigma^2_{\text{Kişi}}$	.208	21.10	.054	8.54
$\sigma^2_{\text{Madde}}$	.014	1.42	.021	3.32
$\sigma^2_{\text{Zaman}}$	.096	9.74	.060	9.49
$\sigma^2_{\text{Kişi*Madde}}$	.109	11.05	.034	5.38
$\sigma^2_{\text{Kişi*Zaman}}$	.210	21.29	.179	28.32
$\sigma^2_{\text{Madde*Zaman}}$	.005	0.51	.005	0.79
$\sigma^2_{\text{Hata}}$	.344	34.89	.279	44.16
<b>Toplam</b>	<b>.986</b>	<b>100</b>	<b>.632</b>	<b>100</b>

ikinci çalışma için  $G_2$  kişilerarası güvenilirlik değerleri daha düşüktür. Kişilerin tüm ölçüm noktalarındaki ortalama puanları üzerinden güvenilirlik katsayısı ( $G_3$ ) ise pozitif duygu için .98, negatif duygu için .96 bulunmuştur.

### Tartışma

PNDÖ'nün Türkçe formunun daha önceki çalışmalarda olumlu ve olumsuz duygulanımdaki kişilerarası farklılıkları anlamak için iyi bir geçerlik ve güvenilirlik gösterdiği bulunmuştur (Gençöz, 2000). Bu çalışmada ise PNDÖ'nün olumlu ve olumsuz duygulanımı ölçerken iyi derecede kişi-içi güvenilirliğe sahip olduğu gösterilmiştir. Olumlu duygulanımı ölçen pozitif duygu alt ölçeğinin kişi-içi değişim güvenilirliği ilk çalışmada .80, ikinci çalışmada .86 olarak bulunmuştur. Olumsuz duygulanımı ölçen negatif duygu alt ölçeğinin ise kişi-içi değişim güvenilirliği ilk çalışmada .79, ikinci çalışmada .87 olarak bulunmuştur. Bu sonuçlar PNDÖ'nün boy-lamsal ve deneyim örnekleme çalışmalarında olumlu ve olumsuz duygulanımı ölçmek amacıyla güvenilir bir şekilde kullanılabileceğini, ölçeğin kişi-içi değişimi iyi düzeyde ölçebildiğini göstermektedir. Bunun yanı sıra, PNDÖ'nün iki çalışmada da tekrarlanan ölçümlerin ortalaması temelinde hesaplanan kişilerarası farkları ( $G_3$ ) oldukça güvenilir şekilde ölçtüğü gösterilmiştir.

Bu bulgular aynı zamanda durumluk ve süreklilik duygunun ayırt edilebilmesinde de PNDÖ'nün güvenilir bir şekilde kullanılabileceğini desteklemektedir. Test-tekrar test katsayısı hesaplanırken kişinin herhangi bir şekilde iki farklı ölçümde farklı puan almasının sistematik olmayan hata payı olarak değerlendirilmesi durumluk değişimlerin sistematik olmayan hata payı olarak yorumlanmasına neden olabilmektedir. Türkçe uyarlama çalışmasında test-tekrar test katsayısı pozitif duygu için .54, negatif duygu için .40 olarak bulunmuştur (Gençöz, 2000). Test-tekrar test katsayısı için kabul edilebilir düzeyin .70 civarı olduğu düşünüldüğünde (Gençöz, 2000), sadece Türkçe uyarlamasında değil PNDÖ'nün orijinal formunun da güvenilirliği ilk bakışta ister istemez sorgulanabilir (bkz. Watson ve ark., 1988). Bu katsayıların düşük çıkmasının sebebi, test-tekrar test katsayısının, kişinin ölçekten farklı zamanlarda alacağı gerçek puanın değişmeyeceği, değişik bir puan alması durumunda ise bunun sistematik olmayan hata payı olacağı varsayımına bağlanabilir (Hu ve ark., 2016). Test-tekrar test katsayısı hesaplanırken katsayının yüksek çıkması için kişinin farklı ölçümlerde aynı ya da birbirine yakın puanlar alması gerekmektedir. Kısa zaman aralıklarında duygu gibi durumluk olarak değişebilen bir kavram ölçülürken "kişinin gerçek puanı değişmez" gibi bir varsayım ile yapılan test-tekrar test katsayısının kabul edilebilir düzeyden düşük çıkması bu açıdan normaldir (Shavelson

ve ark., 1989) ve geleneksel test-tekrar test yaklaşımının da durumluk olarak dalgalanan bir olgunun güvenilirliğini ölçmek için sınırlı kaldığına da işaret etmektedir. Bunun yanı sıra orijinal makale çalışmasında duygunun ölçüldüğü zaman aralığı arttıkça test-tekrar test katsayısının yükseldiği görülmüştür. Örneğin, "Bugün nasıl hissediyorsunuz?" şeklinde yönerge verildiğinde test-tekrar test katsayıları pozitif duygu için .47, negatif duygu için .39 olarak çıkmıştır (Watson ve ark., 1988). Aynı çalışmada yönerge geçen seneyi kapsayacak şekilde verildiğinde test-tekrar test katsayıları pozitif duygu için .63 iken negatif duygu için .60 bulunmuştur (Watson ve ark., 1988). Birinci bahsedilen ölçüm durumluk duyguya yakınken, ikinci bahsedilen ölçüm süreklilik duyguya daha yakındır. Duygulanımın ölçüldüğü zaman dilimi genişledikçe test-tekrar test katsayısının da arttığı görülmektedir. Geniş zaman dilimlerini kapsayarak yapılan duygulanım ölçümleri süreklilik duyguya daha yakın olduğu için yani durumluk değişimlerden nispeten daha az etkileneceği için test-tekrar test katsayısının varsayımlarına daha uygundur ve kabul edilebilir düzeye yakınlaşan sonuçlar vermektedir.

Fakat bu makalede açıkladığımız şekilde, Cranford ve arkadaşlarının (2006) önerdikleri denklemler yoluyla elde edilen güvenilirlik katsayıları, zaman ile gerçekleşen değişimleri de modelleyebildiği için durumluk değişimlere ölçeğin ne kadar duyarlı olduğu hesaplanabilmektedir. Sonuç olarak, PNDÖ'nün durumluk değişimlere karşı da yeterli ve iyi düzeyde duyarlı olduğu görülmüştür.

Birinci ve ikinci çalışmalarda kayıp verilerin olmasının, yani her katılımcının her ölçümde formu doldurmamış olmasının, çalışma sonuçlarının üzerinde bir etkisi olacağı düşünülebilir. Çalışmada kullanılan VARCOMP yöntemi kayıp veriyi dengeleyerek yansız sonuçlar vermektedir (Cranford ve ark., 2006; Shrout ve Lane 2012). Bunun yanı sıra, tekrar eden ölçümlere dayalı çalışmalarda kayıp data olması olağan bir durum olarak görülmektedir (Hu ve ark., 2016; Shrout ve Lane, 2012) ve bu çalışmanın katılım oranları kabul edilebilir düzeydedir.

### Öneriler

Bilimsel araştırmalar yapılırken ölçüm araçlarının geçerli ve güvenilir olması kritik bir öneme sahiptir ve söz konusu olgunun doğru bir şekilde gözlemlenmesi ya da ölçülmesi gerekmektedir (Shrout ve Lane, 2012). Geleneksel göstergeler (örn., Cronbach alfa katsayısı, test-tekrar test katsayısı, vb.) tek ölçüm alınan çalışmalarda kişilerarası alınan ölçümün ne kadar güvenilir olduğunu yansıtmaktadırlar. Fakat bu göstergeler tekrarlanan ölçüm alınan çalışmalarda hem kişilerarası hem de kişi-içi farklılıklardaki güvenilirliği açıklamakta yetersiz kalmaktadırlar. Bu sebeple, özellikle kişi-içi değişimi



ölçmeyi hedefleyen tekrar eden ölçümlere sahip çalışmalarda araştırmacıların geleneksel kişilerarası güvenilirlik göstergelerinin yerine, hem tekrar eden ölçümlere uygun kişilerarası güvenilirlik katsayılarını (özellikle bu makalede belirtilen  $G_3$  katsayısı) hem de kişi-içi değişim güvenilirlik katsayısını rapor etmeleri önerilmektedir. Bununla birlikte, araştırmacıların gelecekte PNDÖ'yü kişi-içi değişimi ölçme amacıyla kullandıkları zaman kişi-içi güvenilirlik katsayılarını da rapor etmelerinin alanyazın açısından faydalı olacağı düşünülmektedir. Ayrıca, birden fazla ölçüm alınan çalışmalarda kişilerarası güvenilirlik rapor edilirken, tek ölçüm alınan çalışmalarda kullanılan kişilerarası güvenilirlik göstergeleri yerine bu makalede yer verilen kişilerarası güvenilirlik göstergelerine (özellikle  $G_3$ 'e) yer verilmesi daha sağlıklı ve betimleyici bilgiler sağlayacaktır. Bu sebeple, birden fazla ölçüm alınan çalışmalarda bu çalışmada belirtilen kişilerarası güvenilirlik göstergelerinin bildirilmesi tavsiye edilmektedir. Araştırmacılar, bu makalenin sonunda verilen SPSS komutlarını kendi verilerine uyarlayarak tekrarlanan ölçümlerin gerek kişi-içi, gerekse kişilerarası güvenilirlik katsayılarını rahatlıkla hesaplayabilirler.

Son olarak, bu çalışma PNDÖ'nün duyulanım gibi birden fazla ölçüm alınarak ölçülen bir kavram için kişilerarası ve kişi-içi farklılıklara duyarlı bir ölçüm aracı olduğunu destekleyerek hem alanyazına hem de klinik uygulama ve değerlendirmelere ışık tutmaktadır. Yürütülen bu çalışmanın sonuçları, PNDÖ'nün birden fazla ölçüm alınan çalışmalarda hem kişi-içi hem de kişilerarası farklılıkları ölçmek için güvenilir bir şekilde kullanılabileceğini göstermektedir.

### Kaynaklar

- Barrett, F. L. ve Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, 74, 967–984. doi: 10.1037/0022-3514.74.4.967
- Becona, E., Vázquez, F. L., Fuentes, M. J. ve del Carmen Lorenzo, M. (1998). Anxiety, affect, depression and cigarette consumption. *Personality and Individual Differences*, 26, 113–119. doi: 10.1016/S0191-8869(98)00129-9
- Bolger, N., Davis, A. ve Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, 54, 579–616. doi: 10.1146/annurev.psych.54.101601.145030
- Bolger, N., De Longis, A., Kessler, R. C. ve Schilling, E. A. (1989). Effects of daily stress on negative mood. *Journal of Personality and Social Psychology*, 57, 808–818. doi: 10.1037/0022-3514.57.5.808
- Brondolo, E., Brady, N., Thompson, S., Tobin, J. N., Casells, A., Sweeney, M., ... Contrada, R. J. (2008). Perceived racism and negative affect: Analyses of trait and state measures of affect in a community sample. *Journal of Social and Clinical Psychology*, 27, 150–173. doi: 10.1521/jscp.2008.27.2.150
- Burger, J. M. ve Caldwell, D. F. (2000). Personality, social activities, job-search behavior and interview success: Distinguishing between PANAS trait positive affect and NEO extraversion. *Motivation and Emotion*, 24, 51–62. doi: 10.1023/A:1005539609679
- Ciarrochi, J. ve Heaven, P. C. L. (2007). Longitudinal examination of the impact of Eysenck's psychoticism dimension on emotional well-being in teenagers. *Personality and Individual Differences*, 42, 597–608. doi: 10.1016/j.paid.2006.07.022
- Cranford, J. A., Shrout, P. E., Iida, M., Rafaeli, E., Yip, T. ve Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably?. *Personality and Social Psychology Bulletin*, 32, 917–929. doi: 10.1177/0146167206287721
- Crawford, J. R. ve Henry, J. D. (2004). The positive and negative affect schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *The British Journal of Clinical Psychology*, 43, 245–65. doi: 10.1348/0144665031752934
- Cronbach, L. J., Gleser, G., Nanda, H. ve Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: John Wiley & Sons.
- Diener, E. ve Emmons, R. A. (1984). The independence of positive and negative affect. *Journal of Personality and Social Psychology*, 47, 1105–1117. doi: 10.1037/0022-3514.47.5.1105
- Eid, M. ve Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology*, 76, 662–676. doi: 10.1037/0022-3514.76.4.662
- Ekman, P. (1984). Expression and the nature of emotion. K. Scherer ve P. Ekman (Ed.), *Approaches to Emotion* içerisinde (319–344). Hillsdale, NJ: Erlbaum.
- Ferreira, J. A., Reitzle, M., Lee, B., Freitas, R. A., Santos, E. R., Alcoforado, L. ve Vondracek, F. W. (2015). Configurations of unemployment, reemployment, and psychological well-being: A longitudinal study of unemployed individuals in Portugal. *Journal of Vocational Behavior*, 91, 54–64. doi: 10.1016/j.jvb.2015.09.004
- Fiske, D. W. ve Rice, L. (1955). Intra-individual response variability. *Psychological Bulletin*, 52, 217. doi: 10.1037/h0045276

- Gençöz, T. (2000). Pozitif ve negatif duygu ölçeği: Geçerlik ve güvenilirlik çalışması. *Türk Psikoloji Dergisi*, 15(46), 19–26.
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10, 255–282. doi: 10.1007/BF02288892
- Hu, Y., Nesselroade, J. R., Erbacher, M. K., Boker, S. M., Burt, S. A., Keel, P. K., ... Klump, K. (2016). Test reliability at the individual level. *Structural Equation Modeling: A Multidisciplinary Journal*, 23, 532–543. doi: 10.1080/10705511.2016.1148605
- Izard, C. E. (1972). *Patterns of emotions: A new analysis of anxiety and depression*. San Diego, CA: Academic Press.
- Izard, C. E. (1972). *Human emotions*. New York: Plenum.
- Jeffery, A. N., Hyland, M. E., Hosking, J. ve Wilkin, T. J. (2014). Mood and its association with metabolic health in adolescents: A longitudinal study, EarlyBird 65. *Pediatric Diabetes*, 15, 599–605. doi: 10.1111/pedi.12125
- Jiang, J., Song, Y., Ke, Y., Wang, R. ve Liu, H. (2016). Is disciplinary culture a moderator between materialism and subjective well-being? A three-wave longitudinal study. *Journal of Happiness Studies*, 17, 1391–1408. doi: 10.1007/s10902-015-9649-1
- Krohne, H. W., Egloff, B., Kohlmann, C. W. ve Tausch, A. (1996). Investigations with a German version of the positive and negative affect schedule (PANAS). *Diagnostica*, 42(2), 139–156.
- Lonigan, C. J., Phillips, B. M. ve Hooe, E. S. (2003). Relations of positive and negative affectivity to anxiety and depression in children: Evidence from a latent variable longitudinal study. *Journal of Consulting and Clinical Psychology*, 71(3), 465.
- Merz, E. L. ve Roesch, S. C. (2011). Modeling trait and state variation using multilevel factor analysis with PANAS daily diary data. *Journal of Research in Personality*, 45, 2–9. doi: 10.1016/j.jrp.2010.11.003
- Mihić, L., Novović, Z., Čolović, P. ve Smederevac, S. (2014). Serbian adaptation of the positive and negative affect schedule (PANAS): Its facets and second-order structure. *Psihologija*, 47, 393–414. doi: 10.2298/PSI1404393M
- Moneta, G. B., Vulpe, A. ve Rogaten, J. (2012). Can positive affect “undo” negative affect? A longitudinal study of affect in studying. *Personality and Individual Differences*, 53, 448–452. doi: 10.1016/j.paid.2012.04.011
- Oinonen, K. A. ve Mazmanian, D. (2001). Effects of oral contraceptives on daily self-ratings of positive and negative affect. *Journal of Psychosomatic Research*, 51, 647–658. doi: 10.1016/S0022-3999(01)00240-9
- Peluso, M. A. M., Savalli, C., Cúri, M., Gorenstein, C. ve Andrade, L. H. (2010). Mood changes in the course of preparation for the Brazilian university admission exam - A longitudinal study. *Revista Brasileira de Psiquiatria*, 32, 30–36. doi: 10.1590/S1516-44462010000100007
- Quale, A. J. ve Schanke, A. K. (2010). Resilience in the face of coping with a severe physical injury: a study of trajectories of adjustment in a rehabilitation setting. *Rehabilitation Psychology*, 55, 12–22. doi: 10.1037/a0018415
- Rabkin, J. G., McElhiney, M., Moran, P., Acree, M. ve Folkman, S. (2009). Depression, distress and positive mood in late-stage cancer: A longitudinal study. *Psycho-Oncology*, 18, 79–86. doi: 10.1002/pon.1386
- Reis, H. T. (1994). Domains of experience: investigating relationship processes from three perspectives. R. Erber ve R. Gilmore (Ed.) *Theoretical Frameworks in Personal Relationships* içerisinde (87–110). Mahwah, NJ: Erlbaum
- Raudenbush, S. W. ve Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Thousand Oaks: Sage.
- Röcke, C., Li, S. ve Smith, J. (2009). Intraindividual variability in positive and negative affect over 45 days: Do older adults fluctuate less than young adults? *Psychology and Aging*, 24, 863–878. doi: 10.1037/a0016276
- Russell, J. A. ve Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of Personality and Social Psychology*, 76, 805–819. doi: 10.1037/0022-3514.76.5.805
- Sandín, B., Chorot, P., Lostao, L., Joiner, T. E., Santed, M. A. ve Valiente, R. M. (1999). Escalas PANAS de afecto positivo y negativo: Validación factorial y convergencia transcultural. *Psicothema*, 11(1), 37–51.
- Schmukle, S. C., Egloff, B. ve Burns, L. R. (2002). The relationship between positive and negative affect in the Positive and Negative Affect Schedule. *Journal of Research in Personality*, 36, 463–475. doi: 10.1016/S0092-6566(02)00007-7
- Shavelson, R. J., Webb, N. M. ve Rowley, G. L. (1989). Generalizability theory. *American Psychologist*, 44(6), 922.
- Shrier, L. A., Shih, M. ve Beardslee, W. R. (2005). Affect and sexual behavior in adolescents: A review of the literature and comparison of momentary sampling with diary and retrospective self-report methods of measurement. *Pediatrics*, 115, 573–582. doi: 10.1542/peds.2004-2073

- Shrout, P. E. ve Lane, S. P. (2012). Psychometrics. M. R. Mehl ve T.S. Conner (Ed.). *Handbook of Research Methods for Studying Daily Life* içerisinde (302–320). New York: Guilford.
- Traub, R. E. (1994). *Reliability for the social sciences: Theory and applications* (Vol. 3). Thousand Oaks: SAGE publications.
- Tellegen, A., Watson, D. ve Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science, 10*, 297–303. doi: 10.1111/1467-9280.00157
- Tugade, M. M. ve Fredrickson, B. L. (2004). Resilient individuals use positive emotions to bounce back from negative emotional experiences. *Journal of Personality and Social Psychology, 86*, 320–333. doi: 10.1037/0022-3514.86.2.320
- Vadiraja, H. S., Rao, M. R., Nagarathna, R., Nagendra, H. R., Rekha, M., Vanitha, N., ... Rao, N. (2009). Effects of yoga program on quality of life and affect in early breast cancer patients undergoing adjuvant radiotherapy: A randomized controlled trial. *Complementary Therapies in Medicine, 17*, 274–280. doi: 10.1016/j.ctim.2009.06.004
- von Kanel, R., Mausbach, B. T., Ancoli-Israel, S., Mills, P. J., Dimsdale, J. E., Patterson, T. L. ve Grant, I. (2014). Positive affect and sleep in spousal alzheimer caregivers: A longitudinal study. *Behavioral Sleep Medicine, 12*, 358–372. doi: 10.1080/15402002.2013.819470
- Watson, D. ve Clark, L. A. (1984). Negative affectivity: the disposition to experience aversive emotional states. *Psychological Bulletin, 96*, 465–490. doi: 10.1037/0033-2909.96.3.465
- Watson, D. ve Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*, 219–235. doi: 10.1037/0033-2909.98.2.219
- Watson, D., Clark, L. A. ve Carey, G. (1988). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology, 97*, 346–353. doi: 10.1037/0021-843X.97.3.346
- Watson, D., Clark, L. A. ve Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–70. doi:10.1037/0022-3514.54.6.1063
- Watson, D., Clark, L. A., McIntyre, C. W. ve Hamaker, S. (1992). Affect, personality, and social activity. *Journal of Personality and Social Psychology, 63*, 1011. doi: 10.1037/0022-3514.63.6.1011
- Watson, D. ve Pennebaker, J. W. (1989). Health complaints, stress, and distress: Exploring the central role of negative affectivity. *Psychological Review, 96*, 234–254. doi: 10.1037/0033-295X.96.2.234
- Watson, D. ve Vaitya, J. (2003). Mood measurement: Current status and future directions. J. A. Schinka, W. F. ve I. B. Weiner (Ed.). *Handbook of Psychology (Vol.2)* içerisinde (351–377). New Jersey: John Wiley & Sons.
- Zautra, A. J., Affleck, G. G., Tennen, H., Reich, J. W. ve Davis, M. C. (2005). Dynamic approaches to emotions and stress in everyday life: Bolger and Zuckerman reloaded with positive as well as negative affects. *Journal of Personality, 73*, 1511–1538. doi: 10.1111/j.0022-3506.2005.00357

**Ek-1.** Varyans bileşenleri için kullanılacak SPSS kodu

1. Öncelikle SPSS dosyasının uzun görünümü (univariate) olmasını sağlamak için aşağıdaki komutu giriniz. (Örnek teşkil etmesi açısından 3 maddeli ve 2 zaman noktasında ölçüm alınmış bir ölçeğe uygulanacak şekilde hazırlanmıştır.)

Geniş görünüm (Multivariate) aşağıdaki gibidir.

Kişi no	Puan1.zaman1	Puan2.zaman1	Puan3.zaman1	Puan1.zaman2	Puan2.zaman2	Puan3.zaman2
1	a	b	c	d	e	f
2	g	h	ı	j	k	l
...	...	...	...	...	...	...

İşleme başlamadan önce gereken uzun görünüm (univariate) aşağıdaki gibidir.

Kişi no	Zaman_no	Madde_no	Puan
1	1	1	a
1	1	2	b
1	1	3	c
1	2	1	d
1	2	2	e
1	2	3	f
2	1	1	g
2	1	2	h
2	1	3	ı
2	2	1	j
2	2	2	k
2	2	3	l
...	...	...	...

SPSS komutu:

VARSTOCASES

/MAKE puan FROM puan1.zaman1 puan2.zaman1 puan3.zaman1 puan1.zaman2 puan2.zaman2 puan3.zaman2

/INDEX = zaman\_no(2) madde\_no(3)

/KEEP = kişi\_no

/NULL = DROP.

2. SPSS dosyasını geniş görünümünden uzun görünümü hale getirdikten sonra aşağıdaki komutu giriniz.

SPSS komutu:

VARCOMP

puan BY kişi\_no zaman\_no madde\_no

/RANDOM = kişi\_no zaman\_no madde\_no

/METHOD = REML

/DESIGN = kişi\_no zaman\_no madde\_no kişi\_no\*zaman\_no kişi\_no\*madde\_no zaman\_no\*madde\_no

/INTERCEPT = INCLUDE.

3. SPSS bu komutlar neticesinde varyans bileşenlerini içeren bir tablo verecektir. Makalede belirtilen denklemlerle kat-sayıları hesaplayabilirsiniz.

## *Summary*

# Within-Person Reliability of Positive and Negative Affect in Repeated Measurements

Ali Can Gök

Middle East Technical University

Emre Selçuk

Middle East Technical University

Tülin Gençöz

Middle East Technical University

Affect is an important construct in understanding psychological functioning utilized as repeated measurements in both scientific research and clinical settings (Merz & Roesch, 2011; Röcke, Li, & Smith, 2009). The Positive and Negative Affect Schedule (PANAS) (Watson, Clark, & Tellegen, 1988) is a widely used instrument in order to measure both positive and negative affect (Schmukle, Egloff, & Burns, 2002; Watson & Vaitya, 2003). In short, the PANAS measures Positive Affect (PA) containing states of being enthusiastic, active, and full of energy. High scores on PA refer to high energy, ability to work with pleasure and concentration. On the other hand, low scores on PA refer to sadness or stagnation. The PANAS measures Negative Affect (NA) containing anger, fear, guilt, and subjective distress. Low scores on NA refer to calmness and tranquility (Watson, et al., 1988; Gençöz 2000).

Repeated measurements are utilized to detect within person changes in affect (Cranford et al., 2006; Eid & Diener, 1999). This enables researcher to track fluctuations in affect while minimizing the retrospective reporting bias (Cranford et al., 2006). However, the detailed picture provided by repeated measures depends on how reliable the instrument is to within person change (Bolger, Davis, & Rafaeli, 2003).

A critical point when dealing with repeated measurements is that variance may stem from two different sources: between and within-person. In cross-sectional design where the researcher takes measurement on one occasion, the variance in the measurement occasion comes from between person differences. The PANAS has been shown to capture between person differences reliably (Watson, et al., 1988; Gençöz 2000). However, in repeated measures, variance stems from not only between person, but also within person differences (Bolger et al., 2003; Cranford et al., 2006; Hu et al., 2016).

According to Fiske and Rice (1955), in order to be able to portion within-person variance, fluctuations in

one person's test scores are to be partitioned from unsystematic measurement error. Thus, differences on one participant's test scores on different measurement occasions should be decomposed into unsystematic measurement error and the systematic change that occurs in the participant (Eid & Diener, 1999). Therefore, as in the case with between-person reliability, within-person reliability is to be examined before proceeding with studies with more than one measurement occasion.

In psychometric or adaptation studies, usually test-retest coefficient is reported. However, test-retest coefficient is simply the correlation coefficient of the same measurement on two different measurement occasions (Guttman, 1945; Hu et al., 2016; Shrout & Lane, 2012; Traub, 1994). In test-retest coefficient, the test score is partitioned as unsystematic measurement error and true score. Moreover, there is an assumption that if one participant scores differently on different measurement occasions, the difference is due to unsystematic measurement error. Hence, it is assumed that the true score of one participant is not changed. Therefore, the correlation between two different measurement occasions indicates how reliable the instrument measures true scores of participants on different time occasions (Shavelson, Webb, & Rowley, 1989). This may be misleading when researchers study within-person change, due to the fact that test-retest coefficient ignores within-person changes (Eid & Diener, 1999; Hu et al., 2016).

To overcome these obstacles, Cranford and his colleagues (2006) developed a solution within the framework of Generalizability Theory (Cronbach, Gleser, Nanda, & Rejaratham, 1972). In their solution, the underlying assumption is that variance comes from multiple sources. Therefore, within person change is systematic and it can be measured (Cranford et al., 2006). This approach allows for computing different indices of between-person reliability (Cranford et al., 2006), which we discuss in the following sections.

An investigation of PANAS's between and within-person reliability will be beneficial for its use in studies utilizing more than one measurement occasions (i.e. longitudinal designs, diary designs), and in studies aiming at differentiating between state and trait affect.

### Method

In the present study, data were drawn from two different sources. First one is the data set used in PANAS's Turkish adaptation study (Gençöz 2000), in which data were gathered on four different measurement occasions (Study 1). The second one is part of the first author's doctoral dissertation study, in which data were gathered on thirteen different measurement occasions (Study 2).

### Participants

**Study 1.** One hundred ninety-nine undergraduate students, 78 male (39.2%), 108 female (54.3%), 13 unspecified (6.5%), participated in the study. Participants' ages ranged between 18 and 34 ( $M = 20.90$ ,  $SD = 1.88$ ). Participants completed the PANAS at least on two measurement occasions. Measurement occasions were three weeks apart.

**Study 2.** In this study, participants were 46 Clinical Psychology graduate students who were performing psychotherapy under supervision. However, the study proceeded with only 41 therapists due to the fact that participants with at least two measurement occasions were included in the study. There were 35 female (85.4%), 6 male (14.6%) participants. Participants' ages ranged between 23 and 29 ( $M = 25.56$ ,  $SD = 1.45$ ). Participants filled out PANAS on thirteen consequent weeks. Measurement occasions were one week apart.

### Instruments

**Positive and Negative Affect Schedule (PANAS).** The PANAS was developed by Watson and his colleagues (1988) in order to measure PA and NA. Its internal consistency coefficient was .88 for PA, and .85 for NA. Test-retest coefficient for the original form is .47. Turkish adaptation study was conducted by Gençöz (2000). Its internal consistency for the Turkish form was .86 for PA, and .85 for NA. Test-retest coefficient for the Turkish form was .47.

### Procedure

As mentioned earlier, within the framework of Generalizability Theory (Cronbach et al., 1972), variance comes from different sources (Shavelson et al., 1989). Variance components may come from person, time, item variance and their interactions (Cranford et al., 2006). For example, for person  $j$ , on item  $i$ , on time  $k$ ,

the score  $P$  can be modeled as:

$$P_{ijk} = \mu + K_j + M_i + Z_k + (KM)_{ij} + (KZ)_{jk} + (MZ)_{ik} + (KMZ)_{ijk} + e_{ijk}$$

Grand mean for all items on all measurement occasions is  $\mu$ ;  $K_j$ ,  $M_i$ ,  $Z_k$ , are effects person, item and time, respectively (Note: Symbols were Turkified for Turkish readers, and are not the same with those on Cranford et al., 2006). The rest of the symbols represent their two and three-way interactions, and  $e_{ijk}$  is the error term (see Cranford et al., 2006). For these effects' variance components, VARCOMP procedure in SPSS can be applied. Formulas for between and within-person reliability indices were derived from Cranford and his colleagues (2006).

### Results

After performing variance components analysis, between and within-person reliability indices were calculated (Cranford et al., 2006). For study 1, within-person reliability for PA was .80, and for NA .79. Between-person reliability for one measurement occasion was .84 for PA, and .81 for NA. Between-person reliability for across measurement occasions, where one participant's score on one measurement occasion was checked against another participant's score on another measurement occasion, was .50 for PA, .47 for NA. Finally, between-person reliability for mean PA and NA across all measurement waves was .95 for PA, and .95 for NA.

Same procedure was performed on study 2. Within-person reliability in study 2 for PA was .86, for NA .87. Between-person reliability for one measurement occasion was .86 for PA, and .67 for NA. Between-person reliability for across measurement occasions, where one participant's score on one measurement occasion was checked against another participant's score on another measurement occasion, was .39 for PA, .18 for NA. Finally, between-person reliability for mean PA and NA across all measurement waves was .98 for PA, and .96 for NA.

### Discussion

The Turkish version of PANAS showed satisfactory between and within-person coefficients in both studies. Moreover, due to its high level of within-person reliability, the PANAS should be considered as an instrument having good discriminative ability in assessing both state and trait affect. We recommend researchers to provide both between and within-person reliability scores when utilizing the PANAS or any other repeated measures in longitudinal and diary studies.