

Ruh Sağlığı Çalışmalarında Klinik Anlamlılık

Volkan AVŞAR¹, Asiye ŞENGÜL AVŞAR²

¹Recep Tayyip Erdoğan Üniversitesi,
Eğitim Fakültesi, Rehberlik ve Psikolojik
Danışmanlık Anabilim Dalı, Rize,
Türkiye

²Recep Tayyip Erdoğan Üniversitesi,
Eğitim Fakültesi, Eğitimde Ölçme ve
Değerlendirme Anabilim Dalı, Rize,
Türkiye

Öz

Ruh sağlığı alanında hizmet veren uzmanlar, bireylerin ruhsal durumlarına ya da yaşadıkları sorunlara yönelik müdahalelerde bulunmaktadır. Bu müdahaleler genellikle terapi, psikoterapi ya da psikolojik danışma adı altında, klinik ortamlarda ve profesyonel bir anlayışla gerçekleştirilmektedir. Klinik olarak yürütülen bu çalışmalarda genellikle, danışanların ön-test ve son-test ölçümleri arasındaki farkın istatistiksel anlamlılığı araştırılmaktadır. Ancak bu farklılık çoğu zaman grup ortalamaları üzerinden hesaplanmakta, bu da bireysel değişimin belirlenmesinde yeterli bulunmamaktadır. Sunulan psikolojik yardım hizmeti sonrasında danışanlarda meydana gelen değişimin klinik olarak anlamlı olup olmadığının değerlendirilmesi önemlidir. Klinik anlamlılık, bireylerin ön-test ile son-test puanlarındaki değişimin klinik açıdan anlamlı olup olmadığını belirlemeye yönelik istatistiksel çalışmalardır. Bu derleme makalesinin amacı; klinik anlamlılığın değerlendirilmesinde yaygın olarak kullanılan çeşitli yöntemleri tanıtmak ve Türkiye’de deneysel çalışmalarda ya da klinik uygulamalarda bireysel test puanlarındaki değişim üzerine odaklanan ruh sağlığı uzmanları veya araştırmacılar için bu yöntemlerin anlaşılmasını kolaylaştırarak kullanımını yaygınlaştırmaktır. Belirlenen amaç doğrultusunda, klinik anlamlılıkla ilgili önemli makaleler incelenmiş ve değerlendirilmiştir. Literatür incelemesinden elde edilen sonuçlara göre, istatistiksel anlamlılık bireylerin klinik anlamda değişiklik gösterdiğini garanti etmemektedir. Özellikle deneysel çalışmalarda ya da klinik ortamlarda yürütülen araştırmalarda bireylerdeki değişimin klinik anlamlılığının değerlendirilip raporlaştırılması gerekli ve önemlidir.

Anahtar Kelimeler: Klinik anlamlılık, ruh sağlığı çalışmaları, psikolojik müdahalelerin etkililiği, bireysel değişim

Abstract

Clinical Significance In Mental Health Studies

Experts who work in the field of mental health, make interventions based on moods or the issues that individuals go through. These interventions are usually carried out under the name of therapy, psychotherapy or psychological counseling in clinical environments with a professional manner. Usually in these clinically conducted studies, the statistical significance of the difference between the pre-test and post-test measurements of the clients is analyzed. However, detection of this difference is mostly made considering the group mean, which falls short in the identification of individual changes. It is critical to analyze the clinical significance of the change in clients following the psychological interventions. Clinical significance is the statistical study to determine whether the differences between the pre-test and post-test scores of individuals are clinically significant. The aim of this review article is to introduce various widely used methods of clinical significance, and to expedite the use of these methods making them easier to understand for mental health experts or researchers who focus on the differences between the individual test scores in experimental studies or clinical practices in Turkey. In accordance with the purpose identified, prominent articles regarding clinical significance were studied and evaluated. Based on results from literature review, statistical significance does not guarantee any clinical changes in individuals. It is necessary and crucial to determinate and report the clinical significance of changes in individuals that attend particularly the experimental studies or studies carried out in clinical environments.

Keywords: Clinical significance, mental health studies, effectiveness of psychological interventions, individual changes

Correspondence / Yazışma:
Volkan AVŞAR, Recep Tayyip Erdoğan
Üniversitesi, Eğitim Fakültesi, Rehberlik
ve Psikolojik Danışmanlık Anabilim
Dalı, Rize, Türkiye

Tel: +90 464 532 84 54 - 2342

E-mail: volkan.avsar@erdogan.edu.tr

Received / Geliş: 22 Şubat 2020

Accepted / Kabul: 10 Mayıs 2020

Online published / Çevrimiçi yayım:
04 Temmuz 2020

©2020 JCBPR, Available online at
<http://www.jcbpr.org/>

Cite this article as: Avşar, V., Şengül Avşar, A. (2021). Ruh Sağlığı Çalışmalarında Klinik Anlamlılık. J Cogn Behav Psychother Res, 10(1), 109-121. <https://doi.org/10.5455/JCBPR.88460>

GİRİŞ

Ruh sağlığı alanında hizmet veren psikiyatrist, psikolog, psikolojik danışman, psikiyatri hemşireleri ya da sosyal hizmet uzmanı gibi uzmanlar yaptıkları uygulamalı çalışmalarda, genellikle bireylerde meydana gelen değişimlere odaklanmaktadır. Klinik olarak yürütülen araştırmalarda bireylerin aldıkları terapi, psikoterapi ya da psikolojik danışma gibi psikolojik yardım hizmetleri sonrasında; yardım almaya başlamadan önceki durumlarına göre bir farklılık, değişim ya da ilerleme göstermeleri beklenmektedir. Bu nedenle danışanlardaki bireysel değişimlerin değerlendirilmesi, danışanların bir bütün olarak ortalamaya göre gelişim göstermelerinden ve gruplar arası karşılaştırmaların istatistiksel açıdan anlamlı olmasından daha önemlidir (Bauer ve ark., 2004; Jabrayilov ve ark., 2016; Jacobson ve ark., 1984; Jacobson ve ark., 1999). Danışanlarda değişimin gerçekleşip gerçekleşmediği ise genellikle psikolojik danışma hizmetinin öncesinde ve sonrasında uygulanan ölçme araçlarından elde edilen puanların istatistiksel karşılaştırılmasıyla belirlenmektedir (Jacobson ve Truax, 1991). Karşılaştırma sonucunda, uygulanan yöntem ya da müdahalenin etkili olup olmadığına karar verilmektedir. Ancak ön-test ve son-test puanlarının karşılaştırılarak var olan farklılıkların istatistiksel açıdan anlamlı olup olmadığı belirlenmesi, grup ortalamaları üzerinden yapılmaktadır. Bu yöntem, bireysel değişimin belirlenmesinde yeterli bulunmamaktadır (Wise, 2004). Dahası istatistiksel açıdan anlamlı bulunan fark, klinik olarak anlamlı bir değişimi göstermeyebilir (Ferguson ve ark., 2002; Hawley, 1995). Bu noktada klinik anlamlılığın değerlendirilmesi önem taşımaktadır.

Klinik anlamlılık; ruh sağlığı uzmanları tarafından yapılan müdahalelerin, danışanların sorunlarıyla başa çıkma da etkili olduğunu gösteren standartların karşılanmasıyla ilgilidir (Jacobson ve Truax, 1991). Klinik anlamlılığın değerlendirilmesine yönelik dikkat çeken öncü çalışmalar Jacobson ve ark. (1984) ile Jacobson ve Truax (1991) tarafından yapılmıştır. Literatürde Jacobson ve Truax (JT) yöntemi olarak bilinen ve sıklıkla raporlaştırılan JT yönteminde, bireysel test puanlarındaki değişimin ya da klinik anlamlılığın (*clinical significance*) değerlendirilmesi “kesme puanı (*cutoff score*)” ve “güvenilir değişim endeksi-GDE (*reliable change index*)” ile yapılmaktadır (Jacobson ve ark., 1984, Truax ve ark., 1991). Bununla birlikte Gulliksen Lord Novick (GLN) (Hsu, 1989), Edwards Nunnally (EN) (Speer, 1992) ve Hageman Arrindell (HA) (Hageman ve Arrindell, 1999) yöntemleri de literatürde yer alan ve klinik anlamlılığın değerlendirilmesinde kullanılan diğer önemli yöntemlerdir.

Bu derleme makalesinin amacı; klinik anlamlılığın değerlendirilmesinde kullanılan bu yöntemleri tanıtmak ve Türkiye’de klinik uygulamalarda bireysel test puanlarındaki değişim üzerine odaklanan ruh sağlığı uzmanları için bu yöntemlerin anlaşılmasında kolaylık sağlayarak kullanımını yaygınlaştırmaktır.

Neden Klinik Anlamlılık?

Klinik olarak anlamlı değişiklik (*clinically significant change*) psikolojik danışma sürecinde danışan ya da hastanın işlevselliklerinde anlamlı bir değişiklik olması anlamına gelmektedir (Bauer ve ark., 2004). Klinik anlamlılıkla ilgili en önemli araştırmalardan biri, Jacobson ve Truax’ın (1991) yaptığı ve 8000’nin üzerinde atıf alan çalışmadır. İlgili çalışmada; klinik anlamlılığın önemi ile istatistiksel anlamlılıktan ve etki büyüklüğünden farkı üzerinde durulmaktadır.

Klinik uygulamaların etkililiğinin, istatistiksel anlamlılık ile değerlendirilmesinin iki önemli sınırlılığı bulunmaktadır (Jacobson ve Truax, 1991). Jacobson ve Truax’a (1991) göre bu sınırlılıklardan ilki; sunulan psikolojik yardım hizmetinden her bir danışanın bireysel olarak etkilenme durumu ile ilgili bilgi vermemesidir. Yapılan müdahale sonucunda danışanların bireysel değişimleriyle ilgili bilgiler ise, müdahalenin etkililiğinin belirlenmesinde son derece önemlidir.

Diğer sınırlılık yapılan müdahalenin etkisinin istatistiksel açıdan anlamlılığının, klinik anlamlılık ile ilgisinin olmamasıdır. Bununla birlikte istatistiksel karşılaştırmalarda raporlanan değerlerden biri de etki büyüklüğüdür. Etki büyüklüğü sadece yapılan çalışmanın etkisinin büyüklüğü ile ilgili bilgi veren ancak klinik anlamlılıkla doğrudan ilişkili olmayan bir istatistiktir (Jacobson ve Truax, 1991; Jacobson ve ark., 1999).

Klinik Anlamlılığın Değerlendirilmesinde Kullanılan Yöntemler

Bu derleme çalışmasında; öncelikle klinik anlamlılığın değerlendirmesinde en temel araştırmalardan biri olan Jacobson ve Truax’ın (1991) çalışması dikkate alınarak JT yöntemi, sonrasında GLN, EN ve HA yöntemleri anlatılmıştır. Ayrıca anlatılan yöntemlerin kolay anlaşılabilmesi amacıyla hipotetik bir örnek üzerinden tüm yöntemler uygulamalı olarak gösterilmiş ve klinik anlamlılığı raporlaştıran bazı çalışmalardan örneklere de yer verilerek klinik anlamlılığın değerlendirilmesine yönelik çeşitli önerilerde bulunulmuştur.

JT Yöntemi

Ruh sağlığı uzmanları yaptıkları müdahalelerin sonucunda danışanlarda meydana gelen değişimi, klinik anlamlılığı değerlendirerek belirlemelidirler (Atkins ve ark., 2005; Jacobson ve Truax, 1991). Klinik anlamlılığın JT yöntemiyle değerlendirilmesinde işlevsel (*functional*) ve işlevsel olmayan (*dysfunctional*) olmak üzere iki grup dikkate alınmaktadır.

İşlevsel grupta, normal özellikteki kişiler yer almaktadır. Bu grupta yer alan kişiler için psikolojik destek almadan sorunlarıyla başa çıkabildikleri söylenebilir. İşlevsel olmayan grupta yer alanlar ise psikolojik desteğe ihtiyacı olan, psikolojik danışma hizmeti alan ya da tedavi olan danışanlar olarak ifade edilebilir. Şekil 1’de işlevsel olan ve işlevsel olmayan gruplar ve kesme puanları (*a*, *b* ve *c*) verilmiştir.

Sunulan psikolojik yardım hizmeti sonunda danışanlarda meydana gelen değişimin JT yöntemine göre klinik anlamlılığının değerlendirilmesi için iki koşulun karşılanması gerekmektedir. Birinci koşulda müdahale öncesinde işlevsel olmayan grupta yer alan danışanların, yapılan müdahale sonrasında işlevsel gruba geçmiş olmaları gerekmektedir. Bu geçişe karar verilirken müdahale öncesinde ve sonrasında kullanılan ölçme aracına ait kesme puanı kullanılmaktadır. Kesme puanı bir ölçme aracının ölçtüğü özellik bakımından, bireyin sahip olması gereken yeterli yetenek seviyesini veya yeterlik seviyesini ifade etmektedir (Livingston ve Zieky, 1982, 1989). Bir başka ifadeyle kesme puanının istenilen yeterlikte olan bireylerle, istenilen yeterlikte olmayan bireylerin ayırt edilmesini sağladığı söylenebilir.

Klinik anlamlılığın değerlendirilmesinde kullanılan ölçme aracına ait belirlenmiş bir kesme puanı bulunmuyorsa, Jacobson ve Truax (1991) tarafından geliştirilen formüller yardımıyla kesme puanları hesaplanabilmektedir. Özetle ilk koşulda, müdahale sonrasında danışanın son test puanının kesme puanı sınırını geçmesi gerekmektedir.

Klinik anlamlılığın değerlendirilmesinde karşılanması gereken ikinci koşul, kesme puanı sınırını geçen danışanlardaki değişimin istatistiksel olarak güvenilir olmasıdır. Bunun için de GDE değerlendirilmektedir. Çalışmanın bu bölümünde sırasıyla JT yöntemine göre farklı durumlarda kesme puanının ve GDE’nin nasıl hesaplanacağına ilişkin bilgiler verilmiştir. Bu kesme puanları Jacobson ve Truax (1991)’a göre aşağıda anlatılmıştır. Ayrıca Tablo 1’de

kesme puanlarının ve GDE’nin hesaplanması için gerekli olan değerler ve bu değerlerin açıklamaları verilmiştir.

Şekil 1’de verilen varyansları eşit, işlevsel olan ve işlevsel olmayan grupların normal dağılımlı olduğu kabul edildiğinde her iki eğrinin ortalamaları (n_1 ve n_3) arasında kalan nokta, *c*, kesme puanıdır ve

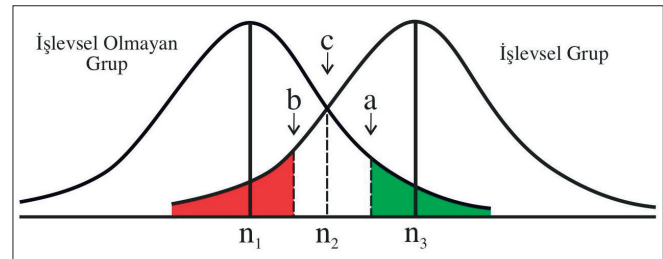
$$c = n_2 = \frac{n_1 + n_3}{2} \quad (1)$$

olarak hesaplanmaktadır. Bu değer, klinik olarak anlamlı değişimin kesme noktasıdır. Yapılan müdahale sonrasında danışanların, klinik olarak anlamlı değişim göstermesi için

Tablo 1: Kesme puanının ve GDE’nin hesaplanması için gerekli olan değerler

Değer	Açıklama
M_1	Deney ve kontrol gruplarının ön-test puan ortalaması
M_2	Deney grubunun (müdahalede bulunulan grubun) son-test puan ortalaması
M_0	İyi işlevsellik gösteren normal örneklemin ortalaması
$s_1 = s_0$	Kontrol grubunun, normal örneklemin ve müdahale öncesi deney grubunun puanlarının standart sapması
s_2	Deney grubunun son-test puanlarının standart sapması
r_{xx}	Ölçümün test tekrar test güvenilirlik katsayısı veya ölçümün iç tutarlık güvenilirlik katsayısı (Cronbach α)
$r_{xx(norm)}$	İyi işlevsellik gösteren normal örneklemden elde edilen güvenilirlik katsayısı
$r_{xx(ön)}$	Müdahale öncesi ön-test puanlarının güvenilirliği
$r_{xx(son)}$	Müdahale sonrası son-test puanlarının güvenilirliği
$r_{xx(fark)}$	Fark puanlarının güvenilirliği
x_1	Herhangi bir bireyin (b_1) ön-test puanı
x_2	Herhangi bir bireyin (b_1) son-test puanı
$r_{ön-son}$	Müdahale öncesi (ön-test) ve müdahale sonrası (son-test) puanları arasındaki korelasyon

Kaynak: Bauer ve ark., 2004, Jacobson ve ark., 1984, Jacobson ve Truax, 1991



Şekil 1. Teorik olarak işlevsel olan ve işlevsel olmayan gruplar ve kesme puanları (Jacobson ve Truax, 1991 s.14).

ilk olarak c noktasını geçmesi gerekmektedir. İşlevsel olan ve işlevsel olmayan grupların varyanslarının eşit olmadığı durumlarda ise kesme puanı c :

$$c = \frac{s_0 M_1 + s_1 M_0}{s_0 + s_1} \quad (2)$$

olarak hesaplanmaktadır. Ruh sağlığı alanında yer alan araştırmalarda, araştırılan değişkenle ilgili olarak normlara ihtiyaç duyulmaktadır ve çoğu zaman bu normlara ulaşılamamaktadır. Böyle durumlarda alternatif kesme puanı a :

$$a = M_1 + 2s_1 \quad (3)$$

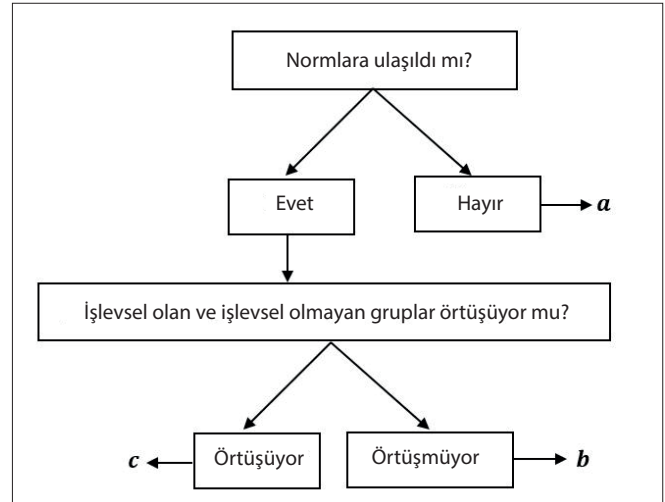
olarak hesaplanmaktadır. İşlevsel olan ve işlevsel olmayan gruplar örtüşmediği (*nonoverlapping*) durumlarda a kesme puanı, yeterince uygun bir ölçüt olmamaktadır. Hesaplamalarda sadece işlevsel grubun değil, aynı zamanda işlevsel olmayan grubun da normlarına ihtiyaç duyulmaktadır. Eğer her araştırmacı a ve c değerlerinin hesaplanmasında kendi çalışmasındaki işlevsel olmayan grup verilerini kullanırsa, bu durumda her çalışmanın farklı bir kesme noktası olacağı için sonuçların karşılaştırabilmesi de mümkün olmayacaktır (Hollon ve Flick, 1988; Wampold ve Jensen, 1986, Akt.: Jacobson ve Truax, 1991). İşlevsel olan ve işlevsel olmayan gruplar örtüşmediğinde b kesme puanı:

$$b = M_0 - 2s_1 \quad (4)$$

olarak hesaplanmaktadır. İşlevsel olan ve işlevsel olmayan gruplar fazla örtüştüğü (*overlapping*) zaman b kesme puanı, a ve c kesme puanlarına göre uyumsuz bir kesme puanı olmaktadır. Ancak bu gruplar örtüşmediğinde b kesme puanı, kullanılması en uygun kesme puanı olarak değerlendirilmektedir. Özetle üç farklı kesme puanından hangisinin daha iyi olduğu, ölçme sonuçlarına ve normlara ulaşıp ulaşılmamasına göre farklılık göstermektedir. Şekil 2'de hangi kesme puanının hangi durumlarda kullanılacağı özetlenmiştir.

Şekil 2'den anlaşılacağı üzere normlara ulaşılmış ise b veya c kesme puanları tercih edilebilir. Normlara ulaşılan durumda işlevsel olan ve işlevsel olmayan gruplar örtüşüyorsa c kesme puanı, aksi durumda b kesme puanı tercih edilmelidir. Normlara ulaşılmadığı durumlarda ise kullanılacak tek kesme puanı, a kesme puanıdır (Jacobson ve Truax, 1991; Jacobson ve ark., 1999).

Hageman ve Arrindell (1999) bireylerin olumsuz duygu durumlarını (kaygı, depresyon ya da somatik belirtiler, psikolojik sıkıntı gibi) belirlemek için kullanılan ölçme araçlarından elde edilen yüksek puanların psikopatolojiye,



Şekil 2. Kesme puanı seçimi.

şikâyetlere veya semptomlara işaret ettiğini; bireylerin olumlu duygu durumlarını (iyi oluş, öfori, neşe, yaşam doyumunu gibi) belirlemek için kullanılan ölçme araçlarından elde edilen yüksek puanların ise iyi işlevselliğe işaret ettiğini ifade etmişlerdir. Ayrıca kullanılan ölçme aracına bağlı olarak JT yönteminde hesaplanan kesme puanlarının değişiklik göstereceğini de belirtmişlerdir. Buna göre 3 ve 4 numaralı eşitlikler, olumlu duygu durumlarını ölçen ölçme araçlarından elde edilen puanlar için hesaplanan kesme puanları olarak tanımlanırken; aşağıda belirtilen 5 ve 6 numaralı eşitlikler, olumsuz duygu durumlarını ölçen ölçme araçlarından elde edilen puanlar için hesaplanan kesme puanları olarak tanımlanmaktadır.

$$a = M_1 - 2s_1 \quad (5) \quad b = M_0 + 2s_1 \quad (6)$$

Kesme puanı, işlevsel olmayan gruptan işlevsel gruba geçiş göstermede bir ölçüt olarak kullanılmaktadır. Özellikle işlevsel olan ve işlevsel olmayan grupların örtüştüğü durumlarda bireylerin ne kadar değişim gösterdiğinin belirlenmesi önemlidir. Böyle durumlarda, işlevsel olmayan grupta yer alan bireyin son-test puanı kesme puanını geçtiğinde, birey işlevsel gruba geçmiş olmaktadır. Ancak söz konusu olan bu değişim istatistiksel olarak anlamlı olmayabilir. Bu değişimin istatistiksel açıdan anlamlı olup olmadığı *GDE* ile belirlenmelidir ve

$$GDE = \frac{x_2 - x_1}{S_{fark}} \quad (7)$$

eşitliğiyle hesaplanmaktadır. Eşitlikte yer alan x_1 bireyin (b_1) ön-test puanını, x_2 bireyin (b_1) son-test puanını, S_{fark} ise iki test puanı arasındaki farkın standart hatasını göstermektedir.

Standart hata ya da ölçmenin standart hatası ($\dot{O}SH$), ölçme sonuçlarının güvenilirliği ile ilgilidir. Eğitim ve psikoloji uygulamalarında en fazla Klasik Test Kuramı'na (KTK) göre ölçekleme yapılmaktadır. KTK'da X gözlenen puan, T gerçek puan ve E hata puanı olmak üzere:

$$X = T + E \quad (8)$$

eşitliğiyle ifade edilmektedir (Crocker ve Algina, 2008). $\dot{O}SH$ bir testten elde edilen gözlenen puanların, gerçek puanlardan sapma derecesinin tahminidir. $\dot{O}SH$, ss standart sapma ve r güvenilirlik katsayısı olmak üzere:

$$\dot{O}SH = ss \sqrt{1 - r} \quad (9)$$

eşitliğiyle ifade edilmektedir (Crocker ve Algina, 2008). $\dot{O}SH$ eşitliğinden anlaşılacağı üzere, ölçme sonuçlarından elde edilen puanların güvenilirliği arttıkça ölçmenin hatası azalacak, bu durumda da gözlenen puanlar gerçek puanlara yaklaşıacaktır. Klinik anlamlılıkta $\dot{O}SH$ 'ın dikkate alınması; yapılan müdahale öncesi ve müdahale sonrası puanlarda meydana gelen değişimin istatistiksel açıdan anlamlı mı yoksa tesadüfi hatalara bağlı olarak mı gerçekleştiğinin belirlenmesinde önemlidir (Ferguson ve ark., 2002). O halde GDE ;

$$S_{fark} = \sqrt{2 (S_{hata})^2} \quad (10) \text{ olmak üzere;}$$

$$GDE = \frac{x_2 - x_1}{S_{fark}} = \frac{x_2 - x_1}{\sqrt{2(S_{hata})^2}} = \frac{x_2 - x_1}{\sqrt{2(s_1 \sqrt{1 - r_{xx}})^2}} \quad (11)$$

olacak şekilde hesaplanmaktadır. Eşitlikte yer alan $\sqrt{2 (S_{hata})^2}$ ifadesindeki $\sqrt{2}$ çarpanı Christensen ve Mendoza (1986) tarafından GDE eşitliğine eklenmiş, diğer bir ifadeyle GDE formülü düzeltilmiştir (Jacobson ve Truax, 1991).

GDE 'nin mutlak değeri; araştırmacı tarafından belirlenen anlamlılık düzeyinde, kritik z puanını aşarsa güvenilir bir değişimin olduğu kabul edilmektedir. Örneğin hesaplanan GDE değerinin; 0,05 anlamlılık düzeyinde $|GDE| \geq 1,96$ veya 0,01 anlamlılık düzeyinde $|GDE| \geq 1,65$ olması bireylerdeki değişimin güvenilir olduğunu, yapılan müdahale sonrası bireyin güvenilir olarak değiştiğini ifade etmektedir (Jabrayilov ve ark., 2016).

Özetle JT yöntemine göre danışanlardaki değişimin klinik anlamlılığının değerlendirilmesinde; öncelikle işlevsel olan ve işlevsel olmayan gruplara göre bir kesme puanı belirlenmektedir. Sonrasında danışanlardaki değişimin şansa bağlı olup olmadığının belirlenmesi için GDE değeri

hesaplanmaktadır (Bauer ve ark., 2004). Diğer bir ifadeyle, danışandaki değişimin klinik olarak anlamlı olduğunu söyleyebilmek için danışanın öncelikle daha önceden belirlenen kesme puanını geçmesi beklenmektedir. Bu değişimin istatistiksel olarak güvenilir olması için GDE 'nin hesaplanması ve değerlendirilmesi gerekmektedir (Anderson ve ark., 2014).

Klinik anlamlılık için kesme puanı ve GDE birlikte değerlendirildiğinde danışanların durumu beş kategoride sınıflandırılabilir (Jabrayilov ve ark., 2016; Jacobson ve Truax, 1991):

1. *İyileşme var (recovery)*: Bu durumda klinik olarak anlamlı değişimin iki koşulu olan kesme puanı ve GDE ölçütlerinin her ikisi de sağlanmaktadır. Diğer bir ifadeyle, danışanda oluşan iyi yönlü değişim, hem istatistiksel hem de klinik açıdan anlamlıdır.
2. *Gelişme var (improvement)*: Bu durumda klinik olarak anlamlı değişimin GDE ölçütü sağlanırken kesme puanı ölçütü sağlanmamaktadır. Diğer bir ifadeyle, danışanda oluşan değişim, istatistiksel açıdan anlamlı ve iyi yönde değişmeyi gösterse de bu değişim klinik olarak anlamlı değildir.
3. *Değişiklik yok (no change)*: Bu durumda klinik olarak anlamlı değişimin iki koşulu olan kesme puanı ve GDE ölçütlerinin her ikisi de sağlanmamaktadır. Diğer bir ifadeyle, danışanda, hem istatistiksel hem de klinik anlamda değişim gerçekleşmemektedir.
4. *Kötüleşme var (deterioration)*: Bu durumda klinik olarak anlamlı değişimin GDE ölçütü sağlanırken kesme puanı ölçütü sağlanmamaktadır. Diğer bir ifadeyle, danışanda oluşan değişim, istatistiksel açıdan anlamlı ve kötü yönde değişmeyi gösterse de bu değişim klinik olarak anlamlı değildir.
5. *Klinik olarak anlamlı kötüleşme var (clinically significant deterioration)*: Bu durumda danışanda oluşan kötü yönlü değişim, hem istatistiksel hem de klinik açıdan anlamlıdır.

JT yöntemine göre yapılan hesaplamalar için ölçme sonuçlarından elde edilen güvenilirlik katsayısına ihtiyaç duyulmaktadır. Bu güvenilirlik katsayısının test tekrar test yönteminden elde edilen güvenilirlik katsayısı mı yoksa iç tutarlık anlamındaki *Cronbach α* güvenilirlik katsayısı mı olacağı tartışılan bir konu olmuştur. Özellikle klinik

örneklerde test tekrar test güvenilirlik katsayısı yerine *Cronbach α* güvenilirlik katsayısının dikkate alınması gerektiği ifade edilmektedir (Martinovich ve ark., 1986; Howard ve ark., 1993; Tingey ve ark., 1996).

JT yöntemine yönelik yapılmış çeşitli eleştiriler bulunmaktadır. Bu eleştirilerin en önemlilerinden biri; işlevsel olan ve işlevsel olmayan örnekler için normlar dikkatli bir biçimde belirlenmediğinde, kesme puanının yapılan çalışmaların örneklemeyle ilgili olarak değişkenlik gösterecek olmasıdır (Bauer ve ark., 2004). Benzer bir diğer eleştiri ise JT'nin karşılaştırmalarda bir standart sağlayamamasıdır. Diğer bir ifadeyle, çalışmalar arasında, araştırılan konu ile ilgili normatif örneklemin tanımlanması ve kullanılması mümkün değildir (Tingey ve ark., 1996). Tüm bu eleştirilere rağmen, hesaplamalarda sağladığı kolaylık nedeniyle klinik anlamlılığın değerlendirilmesinde literatürde en fazla raporlaştırılan yöntemin JT olduğu görülmektedir.

GLN Yöntemi

Bazı durumlarda son-test puanlarındaki değişim, yapılan müdahaleden bağımsız olarak regresyon etkisi nedeniyle gözlenebilir. Ortalamaya doğru regresyon etkisi (regression toward to the mean) ya da regresyon etkisi, tekrarlı ölçümlerde karşılaşılan istatistiksel bir olgudur (Barnett, 2005). İlk kez Galton (1886) tarafından ifade edilmiştir. Galton (1886) çalışmasında 930 çocuk ve ebevninin boylarını ölçmüş, ebevnlerinin boyları ortalamasının üstünde olduğunda çocuklarının boylarının ebevnlerinden kısa; benzer olarak ebevnlerinin boyları ortalamasının altında olduğunda çocuklarının boylarının ebevnlerinden uzun olma eğiliminde olduğunu belirlemiştir. Galton (1886) bu durumu regresyon etkisi olarak açıklamıştır.

Hsu (1995) regresyon etkisini ölçme hatası ile ilişkilendirmiş, her bir bireyin psikolojik danışma sonrası son-test puanlarının, ön-test puanlarına göre ortalamaya daha yakın olacağını ifade etmiştir. Genel bir açıklamayla regresyon etkisi, ilk ölçümleri uç noktalarda olan değişkenlerin, sonraki ölçümlerinin ortalamaya yaklaşmasını ifade etmektedir. Ön-test puanları, normal dağılımdan uzaklaşır (bu durumda uç değerler artar) ve ilgili araştırmalarda kullanılan ölçme araçlarından elde edilen puanların güvenilirliği düşük olursa; regresyon etkisi nedeniyle oluşabilecek iyileşme artış gösterir (Speer, 1992). Hsu (1989) tarafından geliştirilen GLN yönteminde ön-test ve son-test regresyon etkilerinin, son-test puanlarının yorumlanmasında etkili olduğu varsayılmaktadır.

JT yönteminin regresyon etkisini dikkate almaması, GLN yönteminin geliştirilmesinin nedeni olarak belirtilmektedir. Bu yöntemde *GDE* değerinin hesaplamasında, örneklemin ortalaması ve standart sapması dikkate alınmaktadır. GLN yönteminde, evrenin (*population*) ortalama ve standart sapma değerlerine her zaman ulaşamaması ise bu yöntemin en önemli sınırlılığı olarak ifade edilmektedir. Bu durumda daha iyi kestirimlerde bulunabilmek için müdahale öncesi alınan ön-test puanlarına ait ortalamanın *GDE* hesaplamalarında kullanılması önerilmektedir (Atkins ve ark., 2005). Tablo 1'de verilen değerler dikkate alınarak GLN yönteminde evren için ilgili değerlere ulaşıldığında *GDE*:

$$GDE = \frac{(x_2 - M_0) - r_{xx}(x_1 - M_0)}{s_0 \sqrt{1 - r_{xx}^2}} \quad (12)$$

evren için ilgili değerlere ulaşamadığında ise;

$$GDE = \frac{(x_2 - M_1) - r_{xx}(x_1 - M_1)}{s_1 \sqrt{1 - r_{xx}^2}} \quad (13)$$

olacak şekilde hesaplanmaktadır (Atkins ve ark., 2005; Bauer ve ark., 2004).

GLN yöntemine göre hesaplanan *GDE*, JT yönteminde olduğu gibi yorumlanmaktadır.

EN Yöntemi

Speer (1992) tarafından geliştirilen bu yöntemde; eğer regresyon etkisi söz konusu ise ön-test puanlarının güven aralıklarının hesaplanması ve son-test puanlarının da hesaplanan bu güven aralıklarına göre yorumlanması önerilmektedir (Atkins ve ark., 2005). Ancak regresyon etkisi söz konusu değilse, JT yönteminin daha uygun bir yöntem olduğu belirtilmektedir (Speer, 1992). Tablo 1'de verilen değerler dikkate alındığında, EN yönteminde (Atkins ve ark., 2005) *GDE*:

$$P = [r_{xx}(x_1 - M_1) + M_1] \text{ olmak üzere;}$$

$$GDE = P \pm 2s_1 \sqrt{1 - r_{xx}^2} \quad (14)$$

olacak şekilde hesaplanmaktadır. EN yönteminde bireyin son-test puanı, bu yöntemde göre bir aralık olarak hesaplanan *GDE*'nin alt sınırının ($P - 2s_1 \sqrt{1 - r_{xx}^2}$) altında

kalırsa, bireyin güvenilir olarak iyileşmiş olduğu sonucuna ulaşılmaktadır (Atkins ve ark., 2005; Bauer ve ark., 2004).

HA Yöntemi

Hageman ve Arrindell (1999) tarafından geliştirilen bu yöntemde, diğer yöntemlerden farklı olarak klinik olarak anlamlı değişim; bireysel düzeyde (her bir danışan için) ve grup düzeyinde (kesme puanını geçen ve güvenilir değişim gösteren danışanlar için) analiz edilmektedir (Bauer ve ark., 2004). Bu yöntemde klinik anlamlılığın değerlendirilmesi için; bireysel güvenilir değişim endeksi ($GDE_{bireysel} - individual\ reliable\ change$) ve klinik anlamlılık için kesme noktasının bireysel güvenilir geçiş endeksi ($KGE_{bireysel} - the\ index\ for\ individual\ reliable\ passing\ of\ the\ cutoff\ for\ clinical\ significance$) geliştirilmiştir (Hageman ve Arrindell, 1999).

HA yönteminde $GDE_{bireysel}$ hesaplamalarında ön-test ve son-test puanlarının güvenilirliği dikkate alınmaktadır. Ayrıca bu yöntem, regresyon etkisini $GDE_{bireysel}$ hesaplamalarında düzelterek kesme puanını düzenleyen ilk yöntemdir (Atkins ve ark., 2005; Bauer ve ark., 2004). Tablo 1'de verilen değerler dikkate alınarak HA yönteminde $GDE_{bireysel}$ (Hageman ve Arrindell 1999; Bauer ve ark., 2004):

$r_{fark} \geq 0,40$, $S_{örneklem}$: müdahale öncesi standart sapma, $r_{örneklem}$ ölçümün güvenilirliği olmak üzere:

$$S_{hata} = S_{örneklem} \sqrt{1 - r_{örneklem}} \quad (15)$$

$$r_{xx(\text{ön})} = \frac{(S_1^2 - S_{hata}^2)}{S_1^2} \quad (16)$$

$$r_{xx(\text{son})} = \frac{(S_2^2 - S_{hata}^2)}{S_2^2} \quad (17)$$

$$r_{fark} = \frac{s_1^2 r_{xx(\text{ön})} + s_2^2 r_{xx(\text{son})} - 2s_1 s_2 r_{ön-son}}{s_1^2 + s_2^2 - 2s_1 s_2 r_{ön-son}} \quad (18)$$

$$GDE_{bireysel} = \frac{(x_2 - x_1) r_{fark} + (M_2 - M_1) (1 - r_{fark})}{\sqrt{r_{fark}} \sqrt{2 (S_{hata})^2}} \quad (19)$$

olacak şekilde hesaplanmaktadır. $GDE_{bireysel}$ minimum %95 olasılıkla danışanları, bireylerin olumlu duygu durumlarını ölçmeyi amaçlayan ölçme araçları kullanıldığında; " $GDE_{bireysel} < -1,65$ " ise kötüleşme var, " $GDE_{bireysel} \geq -1,65$ " ise güvenilir değişim yok ve " $GDE_{bireysel} > 1,65$ " ise gelişme var şeklinde sınıflandırırken

bireylerin olumsuz duygu durumlarını ölçmeyi amaçlayan ölçme araçları kullanıldığında; " $GDE_{bireysel} > 1,65$ " ise kötüleşme var, " $GDE_{bireysel} \leq 1,65$ " ise güvenilir değişim yok ve " $GDE_{bireysel} < -1,65$ " ise gelişme var şeklinde sınıflandırmaktadır (Hageman ve Arrindell, 1999, s.1174).

$KGE_{bireysel}$ alternatif kesme puanı için gerçek kesme puanı GKP olmak üzere, $KGE_{bireysel}$ değeri ile herhangi bir danışanın belirlenen kesme puanını geçip geçmediğine karar vermek için:

$$KGE_{bireysel} = \frac{M_2 + (x_2 - M_2) r_{xx(\text{son})} - GKP}{(\sqrt{r_{xx(\text{son})}}) S_{hata}} \quad (20)$$

olacak şekilde hesaplanır. Hesaplanan $KGE_{bireysel}$ değerinin yorumlanması iki aşamalıdır. İlk aşamada bireyin son-test puanının minimum %95 olasılıkla doğru sınıflandırıldığı belirlenmesi için $KGE_{bireysel}$ değerinin mutlak değeri 0,05 anlamlılık düzeyinde 1,65 değerinden büyük olmalıdır. İkinci aşamada bireydeki değişimin klinik olarak anlamlı olması için olumlu duygu durumlarını ölçen ölçme araçları kullanıldığında $KGE_{bireysel}$ değerinin 1,65 değerinden büyük, olumsuz duygu durumlarını ölçen ölçme araçları kullanıldığında $KGE_{bireysel}$ değerinin -1,65 değerinden küçük olması gerekmektedir (Hageman ve Arrindell, 1999, s.1176).

$KGE_{bireysel}$ değerinin hesaplanmasında kullanılan GKP , klinik olarak anlamlı değişim için kesme puanıdır. Bu puan, JT yöntemi temel alınarak geliştirilmiştir. JT yönteminde üç farklı kesme puanı olduğu düşünüldüğünde, bu üç farklı kesme puanı için GKP 'de kesme puanları (Hageman ve Arrindell, 1999):

$$a_{GKP} = M_1 + 2S_1 \sqrt{r_{xx(\text{ön})}} \quad (21)$$

$$b_{GKP} = M_0 - 2S_1 \sqrt{r_{xx(\text{ön})}} \quad (22)$$

$$c_{GKP} = \frac{s_0 \sqrt{r_{xx(\text{norm})}} M_1 + s_1 \sqrt{r_{xx(\text{ön})}} M_0}{s_0 \sqrt{r_{xx(\text{norm})}} + s_1 \sqrt{r_{xx(\text{ön})}}} \quad (23)$$

olacak şekilde hesaplanmaktadır. Kullanılan ölçme aracının olumsuz duygu durumlarını ölçmesi durumunda 21 ve 22 numaralı eşitliklerde + ve - değerleri yer değiştirmektedir.

Buraya kadar HA yöntemine göre bireysel düzeyde klinik anlamlılığın nasıl araştırıldığı anlatılmıştır. HA yönteminde diğer yöntemlerden farklı olarak grup düzeyinde de klinik

anlamlılık araştırılmaktadır. Bunun için gerçek fark puanı sıfırdan ($oran_{değişmiş}$) daha küçük (veya büyük) ve gerçek son-test puanı klinik anlamlılık kesme puanından ($oran_{kesme\ puanı\ ötesi}$) daha küçük (veya büyük) olan danışanların oranları (yüzdeleri) kestirilir (Hageman ve Arrindell, 1999). Grup yüzdeleri, normallik varsayımı altında, standart normal dağılım $F(z)$ 'nin kümülatif olasılıkları için tablolara yönlendiren ilgili z puanlarının hesaplanması temelinde elde edilebilir. Buna göre değişen danışanların oranı, z :

\bar{X}_{fark} : ön-test son-test ortalamalarının farkı,

S_{fark} : ön-test son-test fark puanlarının standart sapması,

r_{fark} : ön-test son-test fark puanlarının güvenilirliği olmak üzere:

$$z = \frac{(0 - \bar{X}_{fark})}{S_{fark} \sqrt{r_{fark}}} \quad (24)$$

olarak hesaplanmaktadır. Bireylerin olumsuz duygu durumlarını ölçmek için ölçme araçları kullanıldığında $oran_{değişmiş} = F(z)$, bireylerin olumlu duygu durumlarını ölçmek için ölçme araçları kullanıldığında $oran_{değişmiş} = 1 - F(z)$ değerleri, $F(z)$ tablosundan faydalanarak belirlenmektedir. Kesme puanın ötesinde olan danışanların oranı, z :

GKP : Gerçek kesme puanı,

M_2 : Son test puanlarının ortalaması,

S_2 : Son test puanlarının standart sapması,

$r_{xx(son)}$: Son test puanlarının güvenilirliği olmak üzere:

$$z = \frac{(GKP - M_2)}{(S_2 \sqrt{r_{xx(son)}})} \quad (25)$$

olarak hesaplanmaktadır. Bireylerin olumsuz duygu durumlarını ölçmek için ölçme araçları kullanıldığında $oran_{kesme\ puanı\ ötesi} = F(z)$, bireylerin olumlu duygu durumlarını ölçmek için ölçme araçları kullanıldığında $oran_{kesme\ puanı\ ötesi} = 1 - F(z)$ değerleri, $F(z)$ tablosundan faydalanarak belirlenmektedir. Elde edilen değerler 100 ile çarpılarak grup yüzdeleri elde edilmektedir (Hageman ve Arrindell, 1999).

Bu derleme makalesinde klinik anlamlılık değerlerinin açıklanan yöntemlere göre nasıl hesaplandığı; araştırmacılara ve ruh sağlığı uzmanlarına uygulamalarda yol göstermesi için hipotetik olarak geliştirilen bir örnek üzerinden anlatılmıştır.

Tablo 2: Hipotetik verilerden elde edilen istatistikler

İstatistikler	Değerler
Danışma öncesi örneklem ortalaması	70
Danışma öncesi örneklem ortalamasının standart sapması	25
Danışma sonrası örneklem ortalaması	55
Danışma sonrası örneklem ortalamasının standart sapması	20
Ölçme aracının iç tutarlık güvenilirlik katsayısı	0,90
Ölçmenin standart hatası	$\hat{OSH} = 25 \sqrt{1 - 0,90} = 7,91$
Ön test-son test puanları arasındaki korelasyon	0,65

Bu örnekte bir psikolojik danışmanın, bir grup danışanla sosyal kaygıyı azaltmaya yönelik toplamda 15 hafta süren bireysel psikolojik danışma gerçekleştirdiği varsayılmaktadır. Psikolojik danışman, gruptaki tüm danışanların sosyal kaygı düzeylerini ölçmek amacıyla, danışma öncesinde ve danışma sonrasında, güvenilir ve geçerli sonuçlar veren bir ölçme aracı uygulamıştır. Uygulamalar sonrasında elde edilen puanlara ilişkin Tablo 2'de verilen istatistiklere ulaşılmıştır.

Ön-test puanı 80 ve son-test puanı 50 olan *Danışan A'*'nin, danışma sonrasında sosyal kaygısındaki azalmanın klinik olarak anlamlı olup olmadığı Tablo 2'de verilen bilgilerden yararlanılarak sırasıyla; JT, GLN, EN ve HA yöntemlerine göre açıklanmıştır.

JT Yöntemi ile Klinik Anlamlılığın Değerlendirilmesi

JT yönteminde kesme puanın ve GDE 'nin hesaplanması gerekmektedir. Bu hipotetik örnekte normlara ulaşılmadığı için Şekil 2 ile 5 numaralı eşitlik dikkate alınarak öncelikle kesme puanı hesaplanmalıdır. Buna göre kesme puanı:

$a = M_1 - 2s_1 = 70 - 2 * (25) = 20$ olarak hesaplanır.

Kesme puanı tanımlandıktan sonra GDE , 11 numaralı eşitlik dikkate alınarak hesaplanır.

$$GDE = \frac{x_2 - x_1}{\sqrt{2(S_{hata})^2}} = \frac{50 - 80}{\sqrt{2(7,91)^2}} = \frac{-30}{11,18} = -2,68$$

JT yöntemine göre hesaplanan GDE değeri, 0,05 anlamlılık düzeyinde -1,96 değerinden küçük olmakla birlikte *Danışan A'*'nin son test puanı (50) kesme puanı sınırını

(20) geçemediğinden, bu danışan için gelişme var yorumu yapılır. Danışanda oluşan değişim, iyi yönde gelişme gösterse de bu değişim klinik olarak anlamlı değildir.

GLN Yöntemi ile Klinik Anlamlılığın Değerlendirilmesi

Danışan A'nın GLN yöntemine göre klinik olarak anlamlı değişim gösterip göstermediği bu hipotetik örnekte evrenle ilgili değerlere ulaşamadığı için 13 numaralı eşitlik yardımıyla aşağıdaki gibi hesaplanır:

$$GDE = \frac{(x_2 - M_1) - r_{xx}(x_1 - M_1)}{s_1 \sqrt{1 - r_{xx}^2}}$$

$$= \frac{(50 - 70) - 0,90(80 - 70)}{25 \sqrt{1 - (0,90)^2}} = \frac{-29}{10,90} = -2,66$$

GLN yöntemine göre hesaplanan GDE değeri, 0,05 anlamlılık düzeyinde -1,96 değerinden küçük olmakla birlikte *Danışan A'nın* son test puanı (50), kesme puanı sınırını (20) geçemediğinden, bu danışan için gelişme var yorumu yapılır. Danışanda oluşan değişim, iyi yönde gelişme gösterse de bu değişim klinik olarak anlamlı değildir.

En Yöntemi ile Klinik Anlamlılığın Değerlendirilmesi

EN yöntemi için hipotetik örnekte regresyon etkisinin olduğu varsayılmaktadır. *Danışan A'nın* EN yöntemine göre klinik olarak anlamlı değişim gösterip göstermediği, 14 numaralı eşitlik yardımıyla aşağıdaki gibi hesaplanır:

$$GDE = [r_{xx}(x_1 - M_1) + M_1] \pm 2s_1 \sqrt{1 - r_{xx}^2}$$

$$= [0,90(80 - 70) + 70] \pm 2 * 25 \sqrt{1 - 0,90} = 79 \pm 15,81$$

Buna göre GDE 'nin alt sınırı $79 - 15,81 = 63,19$ 'dur. EN yöntemine göre *Danışan A'nın* son test puanı (50), GDE 'nin alt sınırından (63,19) daha düşük olduğu için danışanda oluşan değişim, güvenilir gelişmeyi göstermektedir.

HA Yöntemi ile Klinik Anlamlılığın Değerlendirilmesi

Danışan A'nın HA yöntemine göre klinik olarak anlamlı değişim gösterip göstermediğinin belirlenmesi için öncelikle $GDE_{bireysel}$ değeri; 16, 17, 18 ve 19 numaralı eşitlikler yardımıyla hesaplanır.

$$r_{xx(\text{ön})} = \frac{(s_1^2 - S_{\text{hata}}^2)}{s_1^2} = \frac{(25^2 - 7,91^2)}{25^2} = 0,90$$

ve

$$r_{xx(\text{son})} = \frac{(s_2^2 - S_{\text{hata}}^2)}{s_2^2} = \frac{(20^2 - 7,91^2)}{20^2} = 0,84$$

$$r_{\text{fark}} = \frac{s_1^2 r_{xx(\text{ön})} + s_2^2 r_{xx(\text{son})} - 2s_1 s_2 r_{\text{ön-son}}}{s_1^2 + s_2^2 - 2s_1 s_2 r_{\text{ön-son}}}$$

$$r_{\text{fark}} = \frac{25^2 * 0,90 + 20^2 * 0,84 - 2 * 25 * 20 * 0,65}{25^2 + 20^2 - 2 * 25 * 20 * 0,65} = \frac{248,50}{375} = 0,66$$

Elde edilen bu değerler, $GDE_{bireysel}$ için 19 numaralı eşitlikte yerine yazıldığında:

$$GDE_{bireysel} = \frac{(x_2 - x_1)r_{\text{fark}} + (M_2 - M_1)(1 - r_{\text{fark}})}{\sqrt{r_{\text{fark}}} \sqrt{2(S_{\text{hata}})^2}}$$

$$= \frac{(50 - 80)0,66 + (55 - 70)(1 - 0,66)}{\sqrt{0,66} \sqrt{2 * (7,91)^2}} = \frac{-24,90}{9,09} = -2,74$$

HA yöntemine göre hesaplanan $GDE_{bireysel}$ değeri, 0,05 anlamlılık düzeyinde -1,65 değerinden küçük olduğu için bireyde güvenilir gelişme var yorumu yapılır. HA yönteminde hesaplanan bir diğer değer ise $KGE_{bireysel}$ değeridir. Bu değer 20 ve 21 numaralı eşitlikler dikkate alınarak aşağıda gösterildiği gibi hesaplanır:

$$a_{GKP} = M_1 - 2s_1 \sqrt{r_{xx(\text{ön})}} = 70 - 2 * 25 \sqrt{0,90} = 22,57$$

$$KGE_{bireysel} = \frac{M_2 + (x_2 - M_2) r_{xx(\text{son})} - GKP}{(\sqrt{r_{xx(\text{son})}}) S_{\text{hata}}}$$

$$= \frac{55 + (50 - 55) * 0,84 - 22,57}{(\sqrt{0,84}) * 7,91} = \frac{28,23}{7,25} = 3,89$$

Hesaplanan $KGE_{bireysel}$ değerinin mutlak değeri, 0,05 anlamlılık düzeyinde 1,65 değerinden büyük olduğu için bireyin son-test puanı minimum %95 olasılıkla doğru sınıflandırılır. Ancak $KGE_{bireysel}$ değeri 0,05 anlamlılık düzeyinde -1,65 değerinden büyük olduğu için *Danışan A'daki* değişim klinik olarak anlamlı değildir.

HA yönteminde bireysel düzeyin yanı sıra grup düzeyinde de klinik anlamlılık araştırılmaktadır. Hipotetik olarak fark puanlarının standart sapmasının 22 olduğu varsayıldığında 24 ve 25 numaralı eşitlikler dikkate alınarak z değeri aşağıdaki gibi hesaplanır.

$$z = \frac{(0 - \bar{X}_{fark})}{S_{fark} \sqrt{r_{fark}}} = \frac{(0 - (55 - 70))}{22 \sqrt{0,66}} = 0,84$$

Hesaplanan değer dikkate alındığında oran_{değişmiş}= $F(z)$ olmak üzere; z tablosundan yararlanılarak oran_{değişmiş}= $F(0,84)=0,7995 \cong 0,80$ elde edilir. Bu değere göre grubun yaklaşık %80'i değişmiştir. Kesme puanı sınırını geçen danışanların oranı, z , Hageman ve Arrindell (1999)'da olduğu gibi GKP için JT yönteminde hesaplanan kesme puanı a dikkate alınarak aşağıdaki gibi hesaplanır.

$$z = \frac{(GKP - M_2)}{(s_2 \sqrt{r_{xx_{son}}})} = \frac{20 - 55}{20 * \sqrt{0,84}} = \frac{-35}{18,33} = -1,91$$

Buna göre oran_{kesme puanı ötesi}= $F(z)$ olmak üzere; z tablosundan yararlanılarak oran_{kesme puanı ötesi}= $F(-1,91)=0,0281 \cong 0,03$ elde edilir. Bu değere göre grup üyelerinin %3'ü kesme puanı sınırını geçmiştir.

Hipotetik olarak verilen bu örnekte klinik anlamlılığın değerlendirilmesinde kullanılan tüm yöntemlere göre *Danışan A*'daki değişim, güvenilir değişim ya da gelişim gösterirken bu değişim ya da gelişim klinik olarak anlamlı bulunmamıştır. Tüm yöntemlerin benzer sonuçlar vermesinin en önemli nedeni yüksek güvenilirlikte (0,90) sonuç veren bir ölçme aracının kullanıldığıdır varsayılmıştır. Atkins ve ark. (2005)'nin çalışmasında yüksek güvenilirlikte sonuç veren ölçme araçları kullanıldığında yöntemlerin birbirleriyle uyumlu oldukları belirtilmiştir.

Klinik Anlamlılıkla İlgili Örnek Araştırmalar

Literatür incelendiğinde psikoloji araştırmalarında klinik anlamlılığı araştıran ve raporlaştıran çok sayıda çalışma olduğu görülmektedir. Örneğin Anderson ve ark. (2014) çalışmalarında çiftler arası uyumun belirlenmesinde kullanılan Çiftler Uyum Ölçeği (Dyadic Adjustment Scale) için kesme puanı ve GDE 'yi hesaplamışlardır. Araştırmalarında istatistiksel olarak grup farklılıklarının değerlendirilmesine rağmen, çiftlerin uyumunda bireysel değişimin klinik olarak anlamlı olup olmadığının belirlenmesi gerektiğini vurgulamışlardır.

Atkins ve ark. (2005) araştırmalarında simülasyon olarak ürettikleri veriler aracılığıyla klinik anlamlılığın değerlendirilmesinde; JT, GLN, EN ve HA yöntemlerini karşılaştırmışlardır. Yapılan araştırmada güvenilirlik katsayılarının ve etki büyüklüklerinin farklı yöntemlere olan etkisi araştırılmıştır. Araştırma sonucunda özellikle yüksek güvenilirlik değerlerinde yöntemlerin birbirleriyle uyumlu oldukları belirlenirken etki büyüklüğü faktörünün yöntemlerin birbirleriyle uyumu üzerinde en az etkisinin olduğu görülmüştür. Araştırmada belirlenen test koşullarında, klinik anlamlılık yöntemleri benzer performans gösterdikleri için birinin diğerine herhangi bir istatistiksel nedenle tercih edilemeyeceği belirtilmiştir. Araştırma sonucunda hem yaygın olarak kullanılması hem de hesaplaması kolay olduğu için klinik anlamlılığın değerlendirilmesinde JT yönteminin tercih edilebileceği, ayrıca örneklere ilişkin bilgilere ulaşılabildiğinde alternatif yöntemlerin daha kesin sonuçlar üreteceği ifade edilmiştir.

McGlinchey ve Jacobson (1999) yaptıkları çalışmada JT yöntemi ile HA yöntemini karşılaştırmışlardır. Karşılaştırma sonucunda her iki yöntem arasında önemli farklılıklar olmadığını belirterek hesaplamalarda kolaylık sağlaması nedeniyle JT yönteminin tercih edilebileceğini ifade etmişlerdir.

Hawley (1995) önleyici müdahale araştırmalarında klinik anlamlılığın değerlendirilmesi üzerine çalışmış ve JT yöntemine göre klinik anlamlılığı belirlemiştir. Bu çalışmanın diğer çalışmalardan farkı, kesme puanının klinik olmayan örneklemelerden elde edilen verilerle hesaplanmış olmasıdır. Hawley (1995), JT yöntemine göre klinik anlamlılığın değerlendirmesinde kesme puanının klinik olmayan örneklemelerden belirlenmesinin zor ve GDE 'nin önleyici müdahalelerin etkililiğinin belirlenmesinde kullanışlı olduğunu belirtmiştir.

Eşkisü (2017) çocuklarda davranış problemlerini azaltmaya yönelik geliştirdiği anababalık destek programının etkililiğini incelediği araştırmasında; ön-test son-test ölçümlü, deney ve kontrol gruplu araştırma desenini kullanmıştır. Araştırmacı, deneysel işlem öncesinde ve sonrasında aldığı ölçümleri istatistiksel olarak karşılaştırmıştır. Bununla birlikte klinik anlamlılığı ve GDE 'yi değerlendirmiştir. Yapılan araştırmada genel olarak geliştirilen programın davranış problemlerini ve dikkat eksikliği ve hiperaktivite bozukluğu belirtilerini klinik olarak anlamlı şekilde azalttığı ifade edilmiştir.

Avşar (2019) Bilişsel Davranışçı Terapi'ye dayalı olarak geliştirdiği 12 oturumluk bireysel psikolojik danışma sürecinin üniversite öğrencilerinin sosyal kaygıları, genel özyeterlikleri ve psikolojik iyi oluşları üzerindeki etkisini değerlendirirken klinik anlamlılıktan yararlanmıştı. Araştırmada yer alan sekiz danışanın tamamının sosyal kaygı ölçümlerinin ön test-son test, ön test-izleme 1 ve ön test-izleme 2 puanları arasında istatistiksel açıdan anlamlı farklılık olduğu belirlenmiştir. Oluşan bu farklılıklar, kesme puanı ve *GDE* koşullarını karşıladığı için klinik olarak anlamlı bulunmuştur. Buna göre danışanların iyi yönlü bir değişim gösterdiği ve yapılan müdahalenin etkili olduğu belirlenmiştir. Bunun yanı sıra, araştırmada alınan genel özyeterlik ve psikolojik iyi oluş ölçümlerindeki değişimler değerlendirildiğinde, danışanların çoğunun ölçümlerinde benzer biçimde istatistiksel açıdan anlamlı farklılık olduğu, kesme puanı ve *GDE* koşullarının karşılandığı görülmüştür. Buna göre, genel olarak, danışanların genel özyeterlik ve psikolojik iyi oluş yönünden de iyi yönlü değişim gösterdikleri belirlenmiştir.

Yukarıda belirtilen çalışmalar özetlendiğinde; klinik anlamlılığın değerlendirilmesi için yeni istatistiksel çözüm arayışlarının olduğu ve buna bağlı olarak yeni yöntemlerin geliştirildiği belirlenmiştir. Ayrıca geliştirilen bu yöntemlerin etkililiği, simülatif veriler aracılığıyla çeşitli test koşulları altında incelenmiştir. Bu yöntemler kullanılarak deneysel araştırmalarda yapılan müdahalelerin etkili olup olmadığının belirlendiği de görülmüştür. Türkiye'de ise klinik anlamlılığın değerlendirildiği sınırlı sayıda araştırmaya ulaşılmıştır.

SONUÇ VE ÖNERİLER

Bireysel test puanlarına ya da bireysel değişime odaklanan ruh sağlığı uzmanları için, yapılan müdahale sonrasında danışanların bireysel değişimlerinin incelenmesi, grup ortalamalarının değerlendirilmesinden daha önemlidir. Bireysel değişimin anlamlılığı ise klinik anlamlılık ile belirlenmektedir. İstatistiksel anlamlılığa ek olarak klinik anlamlılığın değerlendirilmesi, yapılan müdahalenin etkililiğinin ortaya konması açısından önem taşımaktadır. Dolayısıyla ön-test son-test desenli deneysel ya da klinik araştırmalarda klinik anlamlılık değeri dikkate alınarak bireysel değişim değerlendirilmelidir.

Bu derleme çalışması; başta literatürde en fazla atıf alan Jacobson ve Truax (1991)'in araştırması olmak üzere klinik anlamlılığı belirlemede yeni yöntemleri araştıran makaleler gözden geçirilerek oluşturulmuştur. Literatürde klinik

anlamlılığın değerlendirilmesinde en sık raporlaştırılan yöntemin JT yöntemi olduğu, bununla birlikte GLN, EN ve HA yöntemlerine de yer verildiği görülmüştür. Her bir yöntemin birbirinden farklılaşan yönleri bulunmaktadır.

Ruh sağlığı uzmanları; normalara ulaşıp ulaşılmama durumuna, ölçme araçlarından elde edilen puanların güvenilirlik değerlerine, veri setlerinde uç değerlere sahip bireylerin olup olmadığına bağlı olarak klinik anlamlılık için geliştirilen yöntemlerden kendi araştırmaları için en uygun olanını seçmelidirler. Böylece ruh sağlığı uzmanları, seçtikleri yöntemle göre yapılan müdahalenin klinik olarak anlamlılığını kestirerek uyguladıkları yöntem ya da müdahale sonrasında danışanlarda oluşan değişimin klinik olarak anlamlı olup olmadığını belirleyebilirler.

Özellikle Atkins ve ark. (2005) ölçme araçlarından elde edilen puanların güvenilirlik değerleri yüksek olduğunda, klinik anlamlılığın belirlenmesinde kullanılan yöntemlerden ulaşılan sonuçların benzer olduğunu ifade etmişlerdir. Bu bağlamda araştırmalarda, müdahale öncesinde ve sonrasında kullanılan ölçme araçlarının psikometrik özelliklerinin kestirilmesi gerekmektedir. Diğer bir ifadeyle, danışanların doğru değerlendirilebilmesi için geçerli ve güvenilir sonuçlar veren ölçme araçlarının kullanılması gerekli ve önemlidir.

Klinik anlamlılığın araştırılmasında KTK'ya dayalı olarak geliştirilen yöntemlerin sıklıkla çalışıldığı görülmüştür. Bununla birlikte alternatif test kuramlarından Madde Tepki Kuramı'na (MTK) göre yürütülen sınırlı sayıda araştırma (Jabrayilov ve ark., 2016) olduğu belirlenmiştir. Eğitim ve psikoloji alanında uygulama alanı genişleyen MTK uygulamalarına göre ölçekleme yapılarak klinik anlamlılık değerlendirilebilir.

Hakem Değerlendirmesi: Dış bağımsız

Çıkar Çatışması: Yazarların bu araştırma bağlamında açıklaması gereken herhangi bir çıkar çatışması bulunmamaktadır.

Finansal Destek: Bu çalışma için herhangi bir kurumdan finansal destek alınmamıştır.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The authors declare no conflict of interest.

Financial Disclosure: No financial disclosure was received.

KAYNAKLAR

Anderson, S. R., Tambling, R. B., Huff, S. C., Heafner, J., Johnson, L. N. ve Ketring, S. A. (2014). The development of a reliable change index and cutoff for the Revised Dyadic Adjustment Scale. *Journal of Marital and Family Therapy*, 40(4), 525–534. <https://doi.org/10.1111/jmft.12095>

- Atkins, D. C., Bedics, J. D., McGlinchey, J. B. ve Beauchaine, T. P. (2005). Assessing clinical significance: Does it matter which method we use? *Journal of Consulting and Clinical Psychology*, 73(5), 982–989. <https://doi.org/10.1037/0022-006X.73.5.982>
- Avşar, V. (2019). Bilişsel Davranışçı Terapiye dayalı bireysel psikolojik danışmanın sosyal kaygı, özyeterlik ve psikolojik iyi oluş üzerindeki etkisi (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi, Ankara, Türkiye.
- Barnett, A. G., Van Der Pols, J. C. ve Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, 34(1), 215–220. <https://doi.org/10.1093/ije/dyh299>
- Bauer, S., Lambert, M. J. ve Nielsen, S. L. (2004). Clinical significance methods: A comparison of statistical techniques. *Journal of Personality Assessment*, 82(1), 60–70. https://doi.org/10.1207/s15327752jpa8201_11
- Crocker, L. ve Algina, J. (2008). *Introduction to classical and modern test theory*. Mason, Ohio: Cengage Learning.
- Eşkisü, M. (2017). Davranış sorunları olan 3-6 yaş çocuğa sahip anababalar için geliştirilen ‘Anababalık destek programının’ etkililiği (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi, Ankara, Türkiye.
- Ferguson, R. J., Robinson, A. B. ve Splaine, M. (2002). Use of the reliable change index to evaluate clinical significance in SF-36 outcomes. *Quality of Life Research*, 11(6), 509–516. <https://doi.org/10.1023/a:1016350431190>
- Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15, 246–263. <https://doi.org/10.2307/2841583>
- Hageman, W. J. ve Arrindell, W. A. (1999). Establishing clinically significant change: Increment of precision between individual and group level of analysis. *Behaviour Research and Therapy*, 37(12), 1169–1193. [https://doi.org/10.1016/S0005-7967\(99\)00032-7](https://doi.org/10.1016/S0005-7967(99)00032-7)
- Hawley, D. R. (1995). Assessing change with preventive interventions: The reliable change index. *Family Relations*, 44(3), 278–284. <https://doi.org/10.2307/585526>
- Howard, K. I., Lueger, R. J., Maling, M. S. ve Martinovich, Z. (1993). A phase model of psychotherapy outcome: Causal mediation of change. *Journal of Consulting and Clinical Psychology*, 61(4), 678–685. <https://doi.org/10.1037/0022-006x.61.4.678>
- Hsu, L. M. (1989). Reliable changes in psychotherapy: Taking into account regression toward the mean. *Behavioral Assessment*, 11(4), 459–467.
- Hsu, L. M. (1995). Regression toward the mean associated with measurement error and the identification of improvement and deterioration in psychotherapy. *Journal of Consulting and Clinical Psychology*, 63(1), 141–144. <https://doi.org/10.1037/0022-006X.63.1.141>
- Jabrayilov, R., Emons, W. H. ve Sijtsma, K. (2016). Comparison of classical test theory and item response theory in individual change assessment. *Applied Psychological Measurement*, 40(8), 559–572. <https://doi.org/10.1177/0146621616664046>
- Jacobson, N.S., Follette, W.C. ve Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15(4), 336–352. [https://doi.org/10.1016/S0005-7894\(84\)80002-7](https://doi.org/10.1016/S0005-7894(84)80002-7)
- Jacobson, N. S. ve Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12–19. <https://doi.org/10.1037/0022-006X.59.1.12>
- Jacobson, N. S., Roberts, L. J., Berns, S. B. ve McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, 67(3), 300–307. <https://doi.org/10.1037/0022-006x.67.3.300>
- Livingston, S. A. ve Zieky, M. J. (1982). *Passing scores: a manual for setting standards of performance on educational and occupational tests*. New Jersey: Educational Testing Service. https://www.ets.org/Media/Research/pdf/passing_scores.pdf
- Livingston, S. A. ve Zieky, M. J. (1989). A comparative study of standard-setting methods. *Applied Measurement in Education*, 2(2), 121–141. https://doi.org/10.1207/s15324818ame0202_3
- Martinovich, Z., Saunders, S. ve Howard, K. (1996). Some comments on “assessing clinical significance.” *Psychotherapy Research*, 6(2), 124–132. <https://doi.org/10.1080/10503309612331331648>
- McGlinchey, J. B. ve Jacobson, N. S. (1999). Clinically significant but impractical? A response to Hageman and Arrindell. *Behaviour Research and Therapy*, 37(12), 1211–1217. [https://doi.org/10.1016/S0005-7967\(99\)00035-2](https://doi.org/10.1016/S0005-7967(99)00035-2)
- Speer, D. C. (1992) Clinically significant change: Jacobson & Truax (1991) revisited. *Journal of Consulting and Clinical Psychology*, 60(3), 402–408. <https://doi.org/10.1037/0022-006X.60.3.402>
- Tingey, R., Lambert, M. J., Burlingame, G. ve Hansen, N. (1996) Assessing clinical significance: Proposed extensions to method. *Psychotherapy Research*, 6(2), 109–123. <https://doi.org/10.1080/10503309612331331638>
- Wise, E. A. (2004) Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment*, 82(1), 50–59. https://doi.org/10.1207/s15327752jpa8201_10

EXTENDED ENGLISH ABSTRACT

Experts who work in the field of mental health, make interventions based on moods or the issues that individuals go through. These interventions are usually carried out under the name of therapy, psychotherapy or psychological counseling in clinical environments with a professional manner. Usually in these clinically conducted studies, the statistical significance of the difference between the pre-test and post-test measurements of the clients is analyzed. However, detection of this difference is mostly made considering the group mean, which falls short in the identification of individual changes (Wise 2004). Moreover, the statistically meaningful difference may not indicate a clinically significant change (Ferguson et al., 2002; Hawley, 1995). In that sense, it is crucial to evaluate clinical significance. It is critical to analyze the clinical significance of the change in clients following the psychological interventions. Clinical significance is the statistical study to determine whether the differences between the pre-test and post-test scores of individuals are clinically significant.

Clinically significant change means that a significant change occurs in the functionality of the client or the patient during the psychological consultation process (Bauer et al., 2004). One of the most important researches on clinical significance belongs to Jacobson and Truax (1991). In this study, the importance of clinical significance and how it differs from statistical significance as well as effect size is mentioned.

There are two crucial limitations to the evaluation of efficiency of clinical practices alongside statistical significance (Jacobson & Truax, 1991). According to Jacobson and Truax (1991), first of these limitations is that it does not provide any information regarding how each individual client is affected from the psychological services offered while such information about the individual changes of the clients following the intervention is extremely important in identifying the effectiveness of the intervention. The second limitation is that statistical significance of the effectiveness of the intervention is related clinical significance. In addition to this, one of the value reported in statistical comparisons is effect size. Effect size is a statistic that provides information only regarding the effect size of the intervention but is not directly related to clinical significance (Jacobson and Truax, 1991; Jacobson et al., 1999).

Clinical significance is generally determined using particularly Jacobson and Truax (JT) method as well as Gulliksen Lord Novick (GLN), Edwards Nunnally (EN) or Hageman Arrindell (HA) methods. Each method has several different aspects. Since calculations are much easier compared to others, it is seen that the JT method is frequently used as reports in literature.

In the evaluation of clinical significance via the JT method, there are two groups to be considered: functional and dysfunctional. In order to evaluate

the clinical significance of the change that occurs in the clients following the psychological services, two conditions need to be met. The first condition is that clients that were a part of the dysfunctional group before the intervention have to be in the functional group after the intervention. While deciding on this transition, the cutoff from the measurement tool that is utilized before and after the intervention is used. The second condition that needs to be met in the evaluation of clinical significance is that the change in the clients who pass the cutoff is statistically reliable. Reliable Change Index (RCI) is used for this purpose. When the cutoff is evaluated alongside the RCI for clinical significance, the state of the clients is classified in five categories as recovery, improvement, no change, deterioration and clinically significant deterioration (Jabrayilov et al., 2016; Jacobson and Truax, 1991).

In some cases, the change in post-test scores can be observed for the regression effect regardless of the intervention performed. The regression effect towards the mean or the regression effect is a statistical phenomenon encountered in repeated measurements (Barnett, 2005). GLN method should take the regression effect into consideration. The most important limitation of this method is that the population mean and standard deviation values are not always reached. In the EN method, in case of regression effect, the calculation of the confidence intervals of pre-test scores and the interpretation of post-test scores in accordance with the calculated confidence intervals are taken into consideration (Atkins et al., 2005). Unlike other methods, in the HA method, clinically significant change is analyzed on individual level (for each client) and on group level (for clients who pass the cutoff and those who show a reliable change) (Bauer et al., 2004).

The aim of this review article is to introduce these widely used methods of clinical significance, and to expedite the use of these methods making them easier to understand for mental health experts or researchers who focus on the differences between the individual test scores in experimental studies or clinical practices in Turkey.

In accordance with the purpose identified, prominent articles regarding clinical significance were studied and evaluated. Based on results from literature review, statistical significance does not guarantee any clinical changes in individuals. It is necessary and crucial to determinate and report the clinical significance of changes in individuals that attend particularly the experimental studies or studies carried out in clinical environments.

Keywords: clinical significance, mental health studies, effectiveness of psychological interventions, individual changes