



Eğitim Araştırmalarında Güç Analizi ve Bir Uygulama*

Hikmet ŞEVGIN**, Bayram ÇETİN***

Özet: Bu çalışmada eğitim bilimleri çalışmalarında istatistiksel güç analizi konusunda farkındalık yaratmak, alanyazına katkı sağlamak ve eğitim araştırmalarında istatistiksel güç analizinin kullanıldığı uygulamalara yer vermek amaçlanmıştır. Araştırma kapsamında eğitim bilimleri alanında yayın yapmakta olan dergilerden üç tanesi seçkisiz olmayan örnekleme yöntemlerinden ölçüt örnekleme yöntemi ile seçilmiş olup bu dergilerde yer alan toplamda 50 makale incelemeye alınmıştır. Bunlardan 16 tanesi derleme, betimleme türü ve 9 tanesi de ölçek geliştirme-uyarlama türü toplam 25 makale araştırmaya dahil edilmemiştir. Araştırma kapsamında çözümleyici istatistik içeren 25 adet makale incelenmiştir. Betimsel istatistikler ve güç analizi için bazı istatistiksel programlardan yararlanılmıştır. Araştırma kapsamında ilk aşamada eğitim bilimleri alanında yayınlanan makaleler incelenerek eğitim üzerine yapılan araştırmalarda istatistiksel güç analizi konusu irdelenmiştir. Sonrasında gücün hesaplanmasında ihtiyaç duyulan anlamlılık (alfa) seviyesi, etki büyüklüğü ve örneklem büyüklüğü ile ilgili incelemeler yapılmış olup analizlere ilişkin güç hesaplamaları yapılmıştır. Bu inceleme kapsamında sadece bir makalede etki büyüklüğünün hesaplanıp raporlandığı ve analizlere ait istatistiksel gücün ise hesaplanmadığı tespit edilmiştir. Benzer şekilde örneklem büyüklüğü belirlemede de güç analizi yöntemine rastlanmamıştır.

Anahtar Kelimeler: Anlamlılık seviyesi, Etki büyüklüğü, Güç analizi, Örneklem büyüklüğü,

* Bu çalışma IV. Ulusal Eğitimde Ölçme ve Değerlendirme Kongresi'nde sözlü bildiri olarak sunulmuştur.

** Arş. Gör. Gazi Üniversitesi, Gazi Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilimdalı, hikmetsevgin@gmail.com

***Doç. Dr. Gazi Üniversitesi, Gazi Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilimdalı, bcetin27@gmail.com



Power Analysis in Educational Research and an Application

Abstract: In the present study, it was aimed to create awareness about statistical power analysis, contribute to the literature and give place to educational studies using statistical power analysis. In the study, three educational science magazines were selected by criteria sampling method, one of purposeful sampling methods, and a total of 50 articles were studied. A total of 25 articles were not included in the study: 16 of these articles were compilation and description types whereas 9 of these articles were scale development - adaptation types. 25 articles consisting of analytical statistics were examined. For descriptive statistics and power analysis, some statistical soft wares were used. In the first stage of the present study, articles published in the area of educational science were investigated and statistical power subject covered in these articles were studied. Then, significant (alpha) level used in the calculation of the power, effect size and sample size were examined, and power calculations about the analyses were made. As a result, it was determined that in only one essay, effect size was calculated and reported, while statistical power was not calculated. Similarly, power analysis method was not encountered in determining sample size.

KeyWords: Effect Size, Sample Size, Significance (Alpha) Level, Power Analysis

Giriş

Bu araştırmada eğitim bilimleri çalışmalarında istatistiksel güç analizi konusunun önemine dikkat çekmek, alan yazına katkı sağlamak ve eğitim araştırmalarında istatistiksel güç analizinin kullanıldığı uygulamalara yer vermek amaçlanmıştır. Bir istatistiksel gücün hesaplanabilmesi için temelde örneklem büyüklüğü, anlamlılık seviyesi (α) ve etki büyüklüğünün bilinmesi veya belirlenmesi gerekmektedir (Cohen ve diğ., 2003).



Bir araştırma için evreni temsil edecek örneklem sayısı daha planlama aşamasında karar verilmesi gereken bir durumdur. Bir araştırma planlanırken akla gelen ilk soru, ne kadar deneğe ihtiyaç olduğudur (Akgül 2005). Örneklem sayısının ilk baştan belirlenecek olmasının birçok sebebi bulunmaktadır. Çünkü yapılacak araştırma sonucunda ilgili topluluk ya da grup hakkında bir fikir edinme, karar verme, ya da bir çıkarımda bulunma gibi risk taşıyan çok önemli sonuçlara ulaşılmaktadır. Bu sonuçların geçerli ve güvenilir olması bilimsel açıdan büyük bir önem taşımaktadır. Yine yapılacak olan araştırmanın evrene genellenebilmesi için yeterli sayıda örneklem sayısına ihtiyaç duyulmaktadır. Büyüköztürk'e (2011) göre sosyal bilimlerde çoğu araştırma, evrenin tamamı yerine, evreni temsil eden yansız olarak seçilmiş örneklem üzerinden gerçekleşir. Bu durumda araştırmacı, ihtiyaç duyduğu verileri örnekleminden toplar. Bunun da mümkün olması için yeterli sayıda örneklem sayısına ihtiyaç vardır. Aksi takdirde yetersiz bir örneklem sayısı ile yapılacak araştırmanın sonuçlarının yanlış olabileceği gibi, harcanacak olan zaman, emek ve ekonomi açısından istenilmeyen durumların ortaya çıkmasına neden olacaktır.

Evrene genellenebilecek yeterli düzeyde örneklem sayısının belirlenmesinde çeşitli yöntemlerden biri de güç analizidir. Çünkü örneklem büyüklüğü testin gücü ile doğrudan ilişkilidir (Shavelson, 2016). Çalışmanın başlangıcında çalışmanın gücü dikkate alınarak örneklem sayısı belirlenmelidir. Aksi halde çalışmanın gücünde örneklem sayısının yetersizliğinden kaynaklanacak düşüşler, yanlış olabilecek sonuçlara, araştırmacıların emeklerinin-zamanlarının boşa geçmesine ve kaynak israfına neden olacaktır (Cohen ve diğ., 2003; Süt, 2011). Araştırma tasarlanırken seçimlerin çoğu, gücü artırmaya yönelik olarak ele alınır; çünkü istatistiksel gücü düşük olan araştırmalar genelde harcanan çabaya değmeyecektir (Tabachnick G. B, Fidell S. L. Çev. Edt.: Baloğlu M., 2015/2013). Bu nedenle belirli bir güç için ya da yeterli sayıda örneklem için en baştan çalışmaya ait güç analizi



yapılması yerinde bir davranış olacaktır. Çapık (2013) istatistiksel gücün örneklem sayısından etkilendiğini dile getirerek araştırmacıların örneklem sayısını güç analizi ile hesaplayabileceklerini dile getirmektedir. Örneklem sayısının çok az ya da çok fazla seçilmesi durumunda düşülecek belli hatalar gözden kaçırılmaması gereken bir durumdur. Çok küçük bir örneklem üzerinde çalışan bir araştırmacı etki büyüklüğünün büyük olmasına rağmen mevcut olan önemli bir farklılığı tespit edemeyebilir. Yine çok büyük bir örneklem ile yapılan bir çalışmada gerçekte anlamsız olan bir etki yanlış bir şekilde anlamlı olarak çıkabilir (Keskin, 2012; Shavelson, 2016). Çünkü örneklem büyüklüğü örneklem dağılımındaki değişkenliği etkileyerek dolaylı olarak gücü de etkiler (Shavelson, 2016).

Yine örneklem büyüklüğü gibi anlamlılık testi (α) ve etki büyüklüğü de testin gücü ile doğrudan ilişkilidir. Anlamlılık testinde istatistiksel güç, gerçekte yanlış olan sıfır hipotezinin reddedilme olasılığı olarak tanımlanabilmektedir. İstatistiksel güç, araştırılan evrende araştırmacının elde etmek istediği gerçek etkiyi ortaya çıkarabilme oranını ifade eder. Buradaki gerçek etki şansa bağlı olarak elde edilemeyen gerçek değeri belirtmektedir. Anlamlılık seviyesi kurulduğunda ve etki büyüklüğüne bir teori veya yapılan bir ön çalışma ile karar verildiğinde, örneklem sayısı istatistiksel çalışmanın uygulama aşamasından önce belirli bir güç elde etmek için kullanılır. Diğer parametreler sabit iken, örneklem sayısı artarsa güç artmaktadır. Yani örneklem sayısı ile güç arasında doğru orantılı bir ilişki vardır (Keskin, 2012). Araştırmacılar uygun örneklem sayısını, arzu ettikleri gücü, anlamlılık seviyesini ve etki büyüklüğünü kullanarak daha araştırmanın tasarım aşamasında kestirebilmelidir (Shavelson, 2016; Dilullo, 1997).

Hipotez bilimsel araştırmalar neticesinde ortaya çıkan sonuçlara dayalı bir tahminin ifadesidir (Frankel ve Wallen, 2003). Bu tahmini yapmak ve olayları açıklamak için gerçekte iki hipotez kullanılır. Bunlar yokluk hipotezi ve alternatif hipotezdir. Alternatif hipotez

araştırmacının şüphe yönünü ifade eden hipotezdir. Yokluk hipotezi ise araştırmacının iddiasını belirten alternatif hipotezin zıddıdır. (Karagöz ve Ekici, 2004). Şayet araştırmacı iki ortalamanın evren değerlerinin birbirine eşit olmadığını iddia ediyorsa, yokluk hipotezi iki ortalamanın evren değerlerinin birbirine eşit olduğu şeklinde yazılır (Tan, 2016). Araştırmacılar da bulmak veya kanıtlamak istedikleri iddiayı daima alternatif hipotez şeklinde takdim ederler (Karagöz ve Ekici, 2004). Kurulan bir hipotez sonucu dört durum ile karşılaşmamız söz konusudur. Bunlar; Doğru bir hipotezi kabul etme, yanlış bir hipotezi reddetme, yanlış bir hipotezi kabul etme ve doğru bir hipotezi reddetmedir. Yalnız son iki durum istenilen bir durum değildir ve bu durumlar hatalı kararlar alınmasına yol açar. Hipotezle ilgili karar verilirken 2 tip hata yapılır. Bunlar da I. ve II. tip hatalar olarak adlandırılır. Birinci tip hata, yokluk hipotezi doğru olduğu halde red edilmişse oluşur. Böyle bir hatanın yapılma olasılığı α ile gösterilir ve α 'ya testin anlamlılık düzeyi adı verilir. İkinci tip hata ise yokluk hipotezi yanlış olduğu halde red edilememişse oluşur. Böyle bir hatanın yapılma olasılığı β ile gösterilir ve $1 - \beta$ değeri de testin gücü olarak adlandırılır. Dolayısıyla gerçekte doğru olan bir sıfır hipotezinin kabul edilme olasılığı $1 - \alpha$, gerçekte yanlış olan bir sıfır hipotezinin reddedilme olasılığı ise $1 - \beta$ olur (Tan, 2016). Tablo 1 bize hata türlerini ve bunlar arasındaki ilişkiyi daha açıklayıcı bir şekilde vermektedir.

Tablo 1 Hipotezi test etme sürecindeki hatalar

		Gerçek Durum	
		H₀ doğru	H₀ yanlış
Karar	H₀ red edilmedi	Doğru karar $1 - \alpha$	II. tip hata β
	H₀ red edildi	I. tip hata α	Doğru karar $1 - \beta$

Doğru yapılandırılmış bir hipotez, çalışmada kullanılacak istatistiksel analiz yöntemi, çalışmada ele alınan bağımlı ve bağımsız değişkenler konusunda yol göstericidir (Özsoy ve



Özsoy, 2013). Hipotez testi gözlenen bulgunun ortaya çıkma olasılığını vermektedir, ancak bu durum gözlenen bulgunun ya da farkın ne düzeyde olduğuna dair bir bilgi vermemektedir. Işık'a (2014) göre, araştırmacıların esas amacının, ilgilendiği araştırma sorusunun içinde yer alan değişkenler arasında ne tür bağlantıların olduğu ve bu bağlantıların ne kadar güçlü olduğunu anlamak ve açıklamaktır. Bu doğrultuda Işık (2014) veri analizinde yokluk hipotezi anlamlılık testinin sonucunun yanı sıra p değerinin yanında, istatistik test değerinin büyüklüğüne (t , F , r vb.) bakılmasını ve istatistiksel güç (Statistical Power), etki büyüklüğü (Effect Size-ES), güven aralığı (Confidence Interval-CI) gibi ek değerlerin de incelenmesi gerektiğini önermektedir. Yine Büyüköztürk (2011), hipotez testlerinde etki büyüklüğü değerlerinin de hesaplanarak yorumlanmasının sonuçların anlaşılabilirliğini artıracığını belirtmiştir.

Etki büyüklüğü, örneklemden elde edilen sonuçların yokluk hipotezinde tanımlanan beklentilerden sapma düzeyini gösteren istatistiksel değerdir (Cohen, 1988). Yıldırım ve Yıldırım (2011), etki büyüklüğünü yeni denenen bir yöntemin, eskisine kıyasla ne kadar fark yarattığı olarak ifade etmiştir. Murphy ve Myers (2004) değişkenler arasındaki korelasyonel veya kestirimsel ilişkinin incelenilen olgu bakımından ne kadar kuvvetli olduğunu ortaya koyan standart bir ölçü olduğu şekliyle etki büyüklüğünü ifade ettikleri gibi bağımsız değişkenin bağımlı değişken üzerinde ne düzeyde etkiye sahip olduğunu gösteren standart bir ölçü olduğu şekliyle de ifade etmişlerdir. APA; etki büyüklüğü hesaplamalarının pratik anlamlılık için bilimsel araştırmaların raporlanmasında yokluk hipotezi testlerinin desteklenmesi ve daha nitelikli istatistiksel sonuçlar elde edilmesi için önemli bir kriter olduğunu belirtmiş ve araştırmacıların raporlarında etki büyüklüğünü raporlamaları halinde, araştırmaların meta-analiz çalışmalarına dahil edilme şansı yakalayabileceğini, böylelikle alandaki bilgi birikiminin bir parçası haline gelebileceğini belirtmiştir. APA önceden elde



edilen etki büyüklüğü değerleri ile araştırmadan elde edilecek etki büyüklüğü değerlerinin karşılaştırılarak raporlanmasının iyi bir araştırma için olması gereken bir durum olduğuna da değinmiştir. Aşağıdaki Tablo 2’de Cohen tarafından istatistiksel analizlerden bazıları için tavsiye edilen düşük, orta ve yüksek etki büyüklüğü değerleri yer almaktadır.

Tablo 2 Cohen’in Belirlediği Düşük Orta Yüksek Etki Büyüklüğü Değerleri (Cohen, 1988)

Kullanılan Test	Düşük Etki	Orta Etki	Yüksek Etki
T Testi	0,20	0,50	0,80
Varyans Analizi	0,10	0,25	0,40
Korelasyon Analizi	0,10	0,30	0,50
Man-Whitney U testi	0,20	0,50	0,80
Kruskall-Wallis testi	0,10	0,25	0,40
Ki Kare Analizi	0,10	0,30	0,50

Güç, sayısal olarak “0” ile “1” arasında değerler alır. “0” değeri gücün hiç olmadığına “1” değeri ise gücün mükemmel olduğuna işaretler. Bu noktada Cohen (1988), eğer araştırmacının elinde istatistiksel güç seviyesinin ne olması ile ilgili bir dayanağı yoksa bu seviyenin en az 0.80 olması gerektiğini belirtmiştir. Bu seviyede bir güç araştırmacıya H_0 hipotezinin kabul edilip edilmemesi konusunda doğru karar verebilme imkânı sağlayacaktır. Yani araştırmanın gücü %80 ise araştırmacının H_0 hipotezini reddetmek için %80 oranında bir şansa sahip olduğu söylenebilir (Dilullo, 1997).

Araştırmanın Amacı: Bu araştırmanın amacı eğitim bilimleri çalışmalarında istatistiksel güç analizi konusunda farkındalık yaratmak, alanyazına katkı sağlamak ve eğitim araştırmalarında istatistiksel güç analizinin kullanımına yönelik uygulamalara yer vermektir.

Yöntem

Çalışma Grubu: Bu çalışma, betimsel araştırma modelinde tasarlanmıştır. Bu tür araştırmalar, bireylere, gruplara, kurumlara, yöntemlere veya materyallere ait özellikleri, betimleme, karşılaştırma, sınıflama, analiz etme ve yorumlamayı amaçlar (Cohen ve diğerleri, 2011). Araştırmamızın evrenini Türkiye’de eğitim bilimleri alanında yayın yapan dergiler



oluşturmaktadır. Bu dergiler arasından rastgele seçilen üç adet dergi çalışma grubunu oluşturmaktadır. Bu üç dergide yer alan toplamda 50 adet makale incelemeye alınmıştır. Çalışma grubu belirlenirken seçkisiz olmayan örnekleme yöntemlerinden ölçüt örnekleme yöntemi seçilmiştir. Ölçüt olarak tarandığı indeks (SSCI, Alan indeksi, Ulakbim) ve yıl belirlenmiştir. Bu doğrultuda 3 dergi ve bu dergilere ait birer sayı belirlenmiştir. Ölçüt belirlenirken 3 alan uzmanının görüşleri doğrultusunda farklı indekslerde farklı yıllara ait dergilerin son sayıları araştırmaya dâhil edilmiştir. Makalelerin yıllara göre dağılımı ise; birinci derginin 2014 yılına ait son sayısında yer alan 20 makale, ikinci derginin 2015 yılına ait son sayısında yer alan 10 makale ve üçüncü derginin ise 2016 yılı itibari ile en son yayınlanan sayısında yer alan 20 makaleden oluşmaktadır. Veriler Nisan-Mayıs 2016 tarihleri arasında internet aracılığı ile toplanmıştır. Araştırmaya sadece çözümleyici istatistik kullanılan 25 makale dâhil edilmiştir.

Veri Toplama Aracı: Veriler hazırlanan makale tanıtım formu aracılığı ile toplanmıştır. Bu formda analizin adı, örneklem sayısı, anlamlılık seviyesi (α), etki büyüklüğü ve istatistiksel güç hesaplama durumu gibi maddeler yer almaktadır. Araştırmaya makalelerin dâhil edilip edilmeyeceğine karar verme sürecinde istatistiksel anlamlılık testlerinin uygulanıp uygulanmaması etkili olmuştur. Dâhil edilen makalelerde yer alan analizler Cohen'in güç tablolarında ve analizlerimizi gerçekleştirdiğimiz G-Power programında yer alan testler ile gerçekleştirilen analizler (bağımsız gruplarda t testi, χ^2 testi, varyans analizi, Mann-Whitney U testi, Wilcoxon testi ve Kruskal-Wallis testi gibi) araştırmaya dahil edilmiş olup her biri için, Cohen'in (1988) *Statistical Power Analysis for the Behavioral Sciences* adlı kitabından alınan küçük, orta ve büyük seviyedeki etki büyüklüğü değerlerine göre güç analizi hesaplaması yapılmıştır. Makalede belirtilen bir anlamlılık seviyesi olmaması durumunda



istatistiksel gücü hesaplamak için anlamlılık seviyesi 0.05 olarak alınmış olup hipotezin yönü belirtilmediği durumlarda ise hipotezin yönü iki yönlü olarak işleme alınmıştır.

Verilerin Analizi: Yukarıda da değindiğimiz gibi bir istatistiksel gücün hesaplanabilmesi için temelde örneklem büyüklüğü, anlamlılık seviyesi (α) ve etki büyüklüğünün bilinmesi gerekmektedir. Yaptığımız inceleme sonucu eğitim bilimleri alanında yayınlanan dergilerde yer alan makaleler barındırdığı çözümleyici analizler istatistiksel güç açısından incelenmiştir. Makalede belirtilen örneklem sayısı, kurulan hipotezin anlamlılık seviyesi ve etki büyüklüğü işleme alınarak G-Power 3.1 programı üzerinde analizlere ilişkin istatistiksel güç hesaplaması yapılmaya çalışılmıştır. Makalelerde anlamlılık seviyesinin belirtilmediği durumlarda $\alpha=0.05$ anlamlılık düzeyinde hesaplamaya katılmıştır. Makalelerde etki büyüklüğü hesaplamasının sadece bir makalede yapıp raporlandığına rastlanmıştır. Bu çalışmamızda her bir istatistiksel test için Jacop Cohen'in (1988) *Statistical Power Analysis for the Behavioral Sciences* yer verdiği ve her analiz için belirttiği düşük orta ve yüksek etki büyüklüğü sınır değerleri baz alınmıştır.

Üç dergiye ait 50 makale incelemeye alınmış olup bunlardan 16 tanesi derleme, betimleme türü makaleler olduğundan 9 tanesi de ölçek geliştirme-uyarlama türü olduğu için incelemeye alınmamıştır. Kalan 25 makale üzerinden toplamda 53 analizin dergilere dağılımı Tablo 1'de sunulmaktadır.

Tablo 1. Araştırmaya Dâhil Edilen Makalelerin Dergilere ve Analiz Sayılarına Göre Dağılımı

Makale	Sayı	Analiz sayısı	%
A	6	19	35,8
B	10	16	30,2
C	9	18	34,0

Tablo 1’de yer alan makalelere ait 53 analiz için düşük, orta ve yüksek etki büyüklüğü değerlerinde olmak üzere her biri için istatistiksel güç hesaplaması yapılmıştır. Bu hesaplamalara ait verilerin ortalama ve standart sapma değerleri ise Tablo 2’de verilmiştir.

Tablo 2 Kullanılan Testler ve İstatistiksel Güçlerin Dağılımı

Kullanılan Test	n		Düşük Etki Büyükliğünde Güç	Orta Etki Büyükliğünde Güç	Yüksek Etki Büyükliğünde Güç
T Testi	17	Ort.±S.Sap	0,42 ±0,31	0,83 ±0,23	0,94 ±0,11
Varyans Analizi	15	Ort.±S.Sap	0,34 ±0,25	0,78 ±0,32	0,87 ±0,26
Korelasyon Analizi	12	Ort.±S.Sap	0,25 ±0,16	0,86 ±0,23	0,94 ±0,15
Ki Kare Analizi	2	Ort.±S.Sap	0,21 ±0,23	0,54 ±0,62	0,61 ±0,54
Kruskall-Wallis testi	4	Ort.±S.Sap	0,32 ±0,24	0,91 ±0,06	0,99 ±0,05
Mann-Whitney U testi	3	Ort.±S.Sap	0,34 ±0,39	0,63 ±0,38	0,82 ±0,27

Bu çalışmada analizlere ait verileri diğer yayınlar ile kıyaslamak için ortalama, ve standart sapma değerleri verilmiştir. Çalışmada 17’si T testi, 15’i Varyans analizi, 12’si Korelasyon analizi, 2’si Ki-kare analizi, 4’ü Kruskal Wallis testi ve 3’ü Mann-Withney U testi olmak üzere 53 analizin her biri için düşük orta ve yüksek olmak üzere 159 güç analizi hesaplaması yapılmıştır. Tablo 2’de görüldüğü gibi, T testi için düşük etki büyüklüğünde ortalama istatistiksel güç $0,42\pm 0,31$, orta etki büyüklüğünde $0,83\pm 0,23$ ve yüksek etki büyüklüğünde $0,94\pm 0,11$ olarak hesaplanmıştır. Varyans analizi için ortalama istatistiksel güç düşük etki büyüklüğünde $0,34\pm 0,25$, orta etki büyüklüğünde $0,78\pm 0,32$ ve yüksek etki büyüklüğünde $0,94\pm 0,11$ olarak hesaplanmıştır. Korelasyon analizi için ortalama istatistiksel güç düşük etki büyüklüğünde $0,25\pm 0,16$ orta etki büyüklüğünde $0,86\pm 0,23$ ve yüksek etki büyüklüğünde $0,94\pm 0,15$ olarak hesaplanmıştır. Ki kare analizi için ortalama istatistiksel güç düşük etki büyüklüğünde $0,21\pm 0,23$ orta etki büyüklüğünde $0,54\pm 0,62$ ve yüksek etki büyüklüğünde $0,61\pm 0,54$ olarak hesaplanmıştır. Kruskal Wallis testi için ortalama istatistiksel güç düşük etki büyüklüğünde $0,32\pm 0,24$ orta etki büyüklüğünde $0,91\pm 0,06$ ve yüksek etki

büyükliğinde $0,99 \pm 0,05$ olarak hesaplanmıştır. Son olarak Mann Withney U testi için ortalama istatistiksel güç düşük etki büyüklüğünde $0,34 \pm 0,39$ orta etki büyüklüğünde $0,63 \pm 0,38$ ve yüksek etki büyüklüğünde $0,82 \pm 0,27$ olarak hesaplanmıştır.

Araştırmada yer alan analizler parametrik ve non-parametrik özellik taşımadıkları. Bu nedenle incelenen analizler parametrik ve non-parametrik olarak gruplandırılarak incelenmiş ve bu incelemeye ait veriler Tablo 3’de sunulmuştur.

Tablo 3 Analiz Türlerine Göre İstatistiksel Güçlerin Dağılımı

	n		Düşük Etki Büyükliğünde Güç	Orta Etki Büyükliğünde Güç	Yüksek Etki Büyükliğünde Güç
Parametrik	32	Ort. \pm S.Sap	0,38 \pm 0,28	0,81 \pm 0,27	0,91 \pm 0,19
Non Parametrik	21	Ort. \pm S.Sap	0,29 \pm 0,20	0,81 \pm 0,28	0,90 \pm 0,21
Tümü	53	Ort. \pm S.Sap	0,35 \pm 0,26	0,81 \pm 0,27	0,91 \pm 0,20

Parametrik testler için ortalama istatistiksel güç düşük etki büyüklüğünde $0,38 \pm 0,28$ orta etki büyüklüğünde $0,81 \pm 0,27$ ve yüksek etki büyüklüğünde $0,91 \pm 0,19$ olarak hesaplanmıştır. Non-parametrik testler için ortalama istatistiksel güç düşük etki büyüklüğünde $0,29 \pm 0,20$ orta etki büyüklüğünde $0,81 \pm 0,28$ ve yüksek etki büyüklüğünde $0,90 \pm 0,21$ olarak hesaplanmıştır. Parametrik testlerin güçleri düşük etki büyüklüğünde non-parametrik testlere oranla daha yüksek olduğu görülmektedir. Tüm testleri beraber ele aldığımız zaman ortalama istatistiksel güç düşük etki büyüklüğünde $0,35 \pm 0,26$ orta etki büyüklüğünde $0,81 \pm 0,27$ ve yüksek etki büyüklüğünde $0,91 \pm 0,20$ olduğu görülmektedir. Parametrik testlerin güç analizi sonucu elde edilen değerleri non-parametrik testlere göre daha yüksek elde edilmiştir.

Tablo 4 Düşük, Orta, Yüksek Etki Büyüklüğüne Göre İstatistiksel Güçlerin Dağılımı

GÜÇ	Düşük Etki Büyüklüğü			Orta Etki Büyüklüğü			Yüksek Etki Büyüklüğü		
	n	%	‰	n	%	‰	n	%	‰
$\leq 0,25$	24	45,3	45,3	6	11,3	11,3	2	3,8	3,8



0,26 – 0,49	17	32,1	77,4	2	3,8	15,1	1	1,9	5,7
0,50 – 0,79	7	13,2	90,6	8	15,1	30,2	4	7,5	13,2
0,80 – 0,96	3	5,7	96,3	11	20,7	50,9	3	5,7	18,9
0,97 ≤ 1,00	2	3,7	100	26	49,1	100	43	81,1	100

Tablo 4’de araştırma kapsamında incelemeye alınan 53 analizin kendi içinde 0 ile 1 arasında kategorilere ayrıldığı görülmektedir. Araştırmalarda gücün 0,80 ve üzeri olması istenilen bir durumdur. Cohen (1988), eğer araştırmacının elinde istatistiksel güç seviyesinin ne olması ile ilgili bir dayanağı yoksa bu seviyenin en az %80 olması gerektiğini belirtmiştir. Makalelerde görüldüğü gibi analizlerin %90,6’nın istatistiksel gücü düşük bir etkiyi ortaya çıkarabilmek için Cohen’in tavsiye ettiği istatistiksel güç seviyesi olan %80’den daha aşağıdadır. Yine analizlerin %30,2’sinin istatistiksel gücü orta bir etkiyi ortaya çıkarabilmek için gerekli güç seviyesi olan %80’den daha aşağıdır. Son olarak analizlerin %13,2’si istatistiksel gücü yüksek bir etkiyi ortaya çıkarabilmek için gerekli güç seviyesi olan %80’den daha aşağıdır. Düşük etki büyüklüğünde 0,80-0,96 arasında 3 ve 0,97-1,00 arasında ise 2 tane analizin yer aldığı görülmektedir. Orta etki büyüklüğünde ise 0,80-0,96 arasında 11 ve 0,97-1,00 arasında ise 26 tane analizin yer aldığı görülmektedir. Son olarak yüksek etki büyüklüğünde 0,80-0,96 arasında 3 ve 0,97-1,00 arasında ise 43 tane analizin yer aldığı görülmektedir.

Tartışma ve Sonuç

Bu çalışmada eğitim bilimleri alanında yapılan makaleleri inceleyerek eğitim üzerine yapılan araştırmalarda istatistiksel güç analizi konusu irdelendi ve gücün hesaplanmasında ihtiyaç duyulan anlamlılık seviyesi, etki büyüklüğü ve örneklem büyüklüğü ile ilgili incelemeler yapılarak analizlere ilişkin güç hesaplamaları yapıp raporlanmıştır. İnceleme kapsamında sadece bir makalede etki büyüklüğünün hesaplanıp raporlandığı ve diğer



makalelerin hiçbirinde analizlere ait istatistiksel gücün ve etki büyüklüğünün hesaplanmadığı tespit edilmiştir. Bu doğrultuda Çapık'ın da (2013) ifade ettiği gibi ülkemizde istatistiksel gücün yeterince dikkate alınmadığı gibi bir sonuç ta ortaya çıkmaktadır.

Bu çalışmada ülkemizde eğitim bilimleri alanında yayınlanan dergilerden seçkisiz olmayan örnekleme yöntemi ile seçilen dergiler incelenmiş ve inceleme sonucu örneklem büyüklüğünün belirlenmesinde ne tür bir yol izlendiğini anlatan bir rapora ya da yeterli örneklem sayısını verecek bir tabloya, bir formüle herhangi bir şekilde yer verilmediği görülmüştür. Süt (2011) bilimsel çalışmalarda gereğinden az sayıda örneklem incelemenin çalışma sonuçlarının gücünü düşüreceğini, gereğinden fazla örneklem incelemenin ise boş yere emek ve kaynak israfına neden olacağını, bu nedenle çalışmanın başlangıcında çalışma hipotezine uygun olarak gerekli minimum örneklem sayısının belirlenmesiyle çalışma sonuçlarının güvenilirliğinin sağlanmasının yanı sıra kaynak israfının da önüne geçileceğini belirtmiştir.

Bir araştırmaya ait analizin istatistiksel gücünün düşük olması o analizdeki yanlış bir yokluk hipotezinin doğru bir şekilde reddedilme ihtimalinin düşük olması anlamına gelmektedir. Araştırmaya dâhil edilen makalelere ait analizlerin istatistiksel güçlerinin ortalamaları sırasıyla düşük, orta ve yüksek etki büyüklükleri için 0.35, 0.81 ve 0.91 olarak hesaplanmıştır. Bu değerler şu anlama gelir: Şayet bir araştırmacı araştırmasında düşük bir etkiyi kestirmek isteseydi %35 oranında, orta seviyede bir etkiyi kestirmek isteseydi %81 ve yüksek seviyede bir etkiyi kestirmek isteseydi %91 oranında yanlış bir yokluk hipotezini doğru bir şekilde reddedebilme ihtimaline sahip olacaktır.

İncelemeler sonucunda parametrik testlerin non-parametrik testlere oranla özellikle düşük etki büyüklüğü altında daha yüksek oranda güç değerlerine sahip oldukları



görülmektedir. Benzer bir sonuç Çapık'ın (2013) yaptığı sağlık alanındaki bir çalışmasında da görülmektedir. Yine elde ettiğimiz bulgular ışığında düşük etki düzeyinde en yüksek güç değerine 0,42 ortalama değeri ile t-testi analizleri ve en düşük etki düzeyinde en küçük güç değerine ise 0,21 ortalama değeri ile ki-kare analizleri sahiptir. Anlamlılık seviyesinin 0,05 ve etki büyüklüğünün ise Cohen'in önerdiği düşük, orta ve yüksek değerleri sabit olarak aldığımızı düşünürsek ki-kare analizleri için örneklem büyüklüğünün diğer analizlere kıyasla daha düşük olduğu sonucuna varabiliriz.

Cohen (1988), araştırmacıların belirlediği bir güç değerinin olmadığı durumlarda eşik değer için %80 değerinin alınabileceğini tavsiye etmiştir. Görüldüğü gibi analizlerin %90,6'sının düşük etki seviyesinde bu eşik değere ulaşamadığını, %30,2'sinin orta etki seviyesinde bu eşik değere ulaşamadığını ve 13,2'sininde yüksek etki seviyesinde bu eşik değere ulaşamadığını söyleyebiliriz. Araştırmacının araştırmaya konu olan problem durumu, yaygın olarak görülmeyen vakalardan biri ise yani düşük etki büyüklüğü değerine sahipse gerekli olan güç seviyesine ulaşabilmek için gerek anlamlılık seviyesi (0.01, 0.05 ve 0.10 gibi) gerekse yeterli örneklem seviyesinin belirlenmesi araştırma için hayati önem taşımaktadır. Aksi takdirde yanlış sonuçlar ve yanlış yorumlar elde etmek kaçınılmaz olacaktır. Araştırma sonunda elde edilen çıktılar, telafisi mümkün olmayan sonuçları da beraberinde getirebilir. Özellikle sağlık ve psikoloji alanında yapılan araştırmalarda bu yanışa düşme olasılığı daha yüksektir. Araştırmamız kapsamında ele aldığımız makalelerde problem durumu gereği, düşük etki düzeyinde bir etki büyüklüğü altında analiz yapma ihtiyacı hisseden bir araştırmacının gerçek bir sonuca ulaşabilme ihtimali ortalama %9,4 oranında bir değerdir. Buda oldukça düşük bir değerdir.

Araştırmanın planlama aşamasında istenilen güç seviyesine karar vermek daha isabetli bir analiz yapmamıza ve yeterli bir örneklem sayısına ulaşmamıza imkan verecektir. Yine



analiz öncesi hipotezimizin anlamlılık seviyesini (0.01, 0.05 ve 0.10 gibi) problem durumumuzun önem seviyesi ile ilgili olarak belirlememiz, varsa daha önce yapılmış benzer çalışmaları da hesaba katarak etki büyüklüğünü hesaplamamız, daha sağlıklı sonuçlar almamıza olanak verecektir. Yine yapmış olduğumuz bir araştırma sonucu kullandığımız örneklem sayısını, anlamlılık (alfa) seviyemizi ve hesaplamışsa etki büyüklüğünü, hesaplanmamışsa araştırmamızın önemine binaen Cohen'in belirttiği etki seviyelerinden (düşük, orta, yüksek) araştırmamıza uygun olanını hesaba katarak analizlerimize ait istatistiksel gücüde hesaplayabiliriz.

Makalelerde yokluk hipotezi anlamlılık testinin (p) yanı sıra etki büyüklüğü ve istatistiksel güç değerlerinin de raporlanması, yayının kalitesini artıracak gibi meta analiz çalışmalarında kıyaslanmasına da olanak sağlayacaktır. Işık (2014), dergilerin ve editörlerinin p değerine ait etki büyüklüklerinin raporlanmasını zorunlu hale getirmesi araştırmacıların etki büyüklüğü hesaplamalarını araştırma sürecine entegre etmeleri için önemli bir rol oynayacağını ifade etmiştir. Benzer şekilde APA yokluk hipotezi anlamlılık testi (Null Hypothesis Significance Testing – NHST) tartışmalarının sonucunda, araştırmaların bilimsel katkısını arttırmak için bulguların raporlanmasında esas alınacak ilkeleri zaman içinde değiştirmiş; örneğin, etki büyüklüğünü raporlama gereği altıncı basımda (2010) kesinlik kazanmıştır (Akt: Işık, 2014).

Yapılan bir araştırmada 0.05 anlamlılık seviyesinde manidar çıkmayan fark 0.06 alındığında manidar çıkabilir. Buradaki 0.01'lik farkın araştırmamız için ne kadarlık bir öneme sahip olduğuna güç analizi yöntemi ile de bakabiliriz. Analiz sırasında 0.05 ile 0.06 anlamlılık seviyelerini teker teker hesaplayarak elde edilen sonuçlar arasındaki güç farkının araştırmamız için önemine karar verebiliriz. Tan'ın (2016) da ifade ettiği gibi araştırmalarda sadece hipotezlerden hangisinin kabul edildiğinin rapor edilmesi yeterli olmayabilir. Ulaşılan



sonucun güven aralığı bulunup yorumlanabilir. Manidar bir farklılıktan sonra yapılması gereken diğer bir işlem de sonucun teoriye uygunluğu, açıklanan varyans oranı veya etki büyüklüğüne yönelik pratikteki anlamlılığını belirten ilave açıklama veya yorumlarında raporlanması gerekmektedir.

Öneriler

- Araştırmacıların planlama aşamasında araştırma için gerekli örneklem sayısını belirlemede güç analizi yönteminden faydalanabilirler.
- Ya da gerekli verilerin toplanıp analizleri yapılmış bir araştırmanın istatistiksel gücü hesaplanıp raporlanabilir.
- Güç analizi elle hesaplanabileceği gibi gerekli hesaplamaları yapan programlardan da (G-Power, PASS, STATISTICA, NCSS ve Power and Sample gibi) yararlanılabilir.
- Araştırmalarda raporlanan yokluk hipotezi anlamlılık testinin (p) yanında istatistiksel güç analizine ait değer ve etki büyüklüğüne ait değer de raporlanabilir.

Makalenin Bilimdeki Konumu (Yeri)

Eğitim Bilimleri Bölümü/ Eğitimde Ölçme ve Değerlendirme Anabilim Dalı

Makalenin Bilimdeki Özgünlüğü

Eğitim Bilimleri alanında güç analizi adına farkındalık yaratmak ve yapılan çalışmalarda anlamlılık testinin (p) raporlanmasının yanında etki büyüklüğü ile analizin gücünün de raporlanmasının önemine değinmek özgünlüğüne işaret eder.

Kaynakça



Akgül A., 2005, *Tıbbi Araştırmalarda İstatistiksel Analiz Teknikleri SPSS Uygulamaları (3. Baskı)*, Emek Ofset, Ankara.

American Psychological Association. (2010). *Publication manual of the American Psychological Association (6. Baskı)*. Washington, DC.

Büyüköztürk Ş., 2011, *Sosyal Bilimler İçin Veri Analizi El Kitabı (14. Baskı)*. Ankara: PegemA Yayıncılık.

Cohen J. (1988), *Statistical power analysis for the Behavioral Sciences (3. Baskı)*. Lawrence Erlbaum Associates, New Jersey.

Cohen J., Cohen P., West S. G., Aiken L. S., (2003) *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, Third Edition*, Lawrence Erlbaum Associates, Publishers, London.

Cohen, L., Manion, L. ve Morrison, K. (2011). *Research Methods İn Education. (7. Baskı)*, Routledge Falmer Publications, London.

Çapık C., (2013), Bir Hemşirelik Dergisinde Yayınlanan Makalelerde İstatistiksel Güçlerin İncelenmesi, *Anadolu Hemşirelik ve Sağlık Bilimleri Dergisi*, 2013;16:3

Dıllullo, L. K. (1997), A Post Hoc Power Analysis of Inferential Research Examining The Relationship Between Mathematic Anxiety and Mathematic Performance. (Yayınlanmış doktora tezi), Auburn University, Alabama, USA.

Frankel J. R; Wallen, N. E; (2003). *How to Design and Evaluate Research in Education*, McGraw – Hill Education, New York.



Karagöz Y, Ekici S.,(2004), Sosyal Bilimlerde Yapılan Uygulamalı Araştırmalarda Kullanılan İstatistiksel Teknikler ve Ölçekler, *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, Cilt 5, Sayı 1.

Keskin B., (2012), İstatistiksel Güç Analizi: Sosyal Bilimler Alanında Bir Uygulama, (Yayınlanmış Yüksek Lisans Tezi). Akdeniz Üniversitesi, Sosyal Bilimler Enstitüsü, Antalya.

Işık, İ., (2014), Yokluk Hipotezi Anlamlılık Testi ve Etki Büyüklüğü Tartışmalarının Psikoloji Araştırmalarına Yansımaları, *Eleştirel Psikoloji Bülteni* 5. Sayı.

Murphy, K., Myers, B. ve Wolach, A. (2004), *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*, (2. Baskı), Routledge publishing, New York.

Shavelson R. J., (2016), *Sosyal Bilimler için İstatistik 3. Baskı - Statistic for Social Sciences, 3. Edition*, (Çev. Edt. Güler N.), Pegem Akademi Yayıncılık, Ankara.

Süt, N, 2011, Klinik araştırmalarda örneklem sayısının belirlenmesi ve güç (power) analizi, *RAED Dergisi* 2011;3(1-2):29-33 doi:10.2399/raed.11.005.

Tabachnick G. B, Fidell S. L. (2015) *Çok Değişkenli İstatistiklerin Kullanımı - Using Multivariate Statistics (6. Baskı)*, (Çev. Edt., Baloğlu, M.), Ankara: Nobel Akademik Yayıncılık, Ankara.

Tan Ş., (2016), *SPSS ve Excel Uygulamalı Temel İstatistik – 1 (1. Baskı)*, Pegem Akademi Yayıncılık, Ankara.



Özsoy, S., Özsoy, G., (2013), Eğitim Araştırmalarında Etki Büyüklüğü Raporlanması. *İlköğretim Online*, Sayı 1, ISSN 1305-3515.

Yıldırım, H. H. ve Yıldırım, S., (2011), Hipotez Testi, Güven Aralığı, Etki Büyüklüğü ve Merkezi Olmayan Olasılık Dağılımları Üzerine, *İlköğretim Online*, 10(3), 1112-1123.