

## KAĞIT-KALEM VE BİLGİSAYAR FORMUNDA UYGULANAN TESTLERDE ÖĞRENCİ BAŞARISININ KARŞILAŞTIRILDIĞI ÇALIŞMALARIN META-ANALİZİ\*

### META-ANALYSIS OF COMPARATIVE STUDIES OF STUDENT ACHIEVEMENT ON PAPER-PENCIL AND COMPUTER-BASED TEST

Eren Can AYBEK\*, Döndü Bahar ŞAHİN\*\*, Hayriye Merve ERİŞ\*\*\*, Ahmet Salih ŞİMŞEK\*\*\*\*, Mustafa KÖSE\*\*\*\*\*

**ÖZET:** Bilgisayar kullanılarak uygulanan testlerle geleneksel kağıt-kalem testlerini karşılaştıran bir çok araştırma bulunmaktadır. Bu araştırmalar çok çeşitli alanlara, farklı değişkenlere, farklı çalışma gruplarına yayılmıştır. Bu çalışmaların sonuçlarının meta-analiz yardımıyla toparlanıp derlenmesi hem diğer araştırmacılar hem de karar alıcı yetkililer açısından önemlidir. Bu araştırmada, kağıt-kalem formu ve bilgisayar ortamında uygulanan testlerde gösterilen öğrenci performanslarını karşılaştıran çalışmalar taranmıştır. Bu çalışmalar üzerinde kağıt-kalem formu ve bilgisayar ortamında uygulanan testlerin öğrenci performansı bakımından karşılaştırılması ile ilgili meta analiz çalışması yapılmıştır. Meta analize dahil edilen çalışmaların yanlı olup olmadığını belirlemek için duyarlılık analizi olarak Funnel grafiği ve Begg-Mazumdar sıra korelasyonları testinden yararlanılmıştır. İncelenen çalışmalara ait etki büyüklükleri Comprehensive Meta Analysis yazılımı ile incelenmiştir. Hangi etki modelinin benimseneceğine karar vermek amacıyla heterojenlik testi sonuçları incelenmiş ve çalışmalara ait etki büyüklüklerinin normal dağılım gösterdiği görülmüştür. Bu bağlamda sabit etki modelinin seçilmesinde bir sakınca görülmemiştir. Ortak etki büyüklüğü hesaplamasında Hedge'in g katsayısının kullanılması uygun görülmüştür. Çalışmanın iki temel sınırlılığı söz konusudur. Zamansal sınırlılıklar nedeniyle, kağıt-kalem formunda ve bilgisayar ortamında uygulanan testlere ait karşılaştırmaların yapıldığı çalışmaların tümüne ulaşmak mümkün olmamıştır. İkinci sınırlılık ise, meta-analiz sonuçları konu alanlarına göre karşılaştırılmamış olmasıdır. Gelecek araştırmalarda konu alanlarına göre meta-analiz sonuçlarının karşılaştırılması faydalı olacaktır. Meta-analize dahil edilen 9 farklı çalışmaya ait toplam 35 bulguya göre, 29 karşılaştırmada bilgisayar ve kağıt-kalem formları arasında manidar bir fark bulunmazken, 6 çalışmada bilgisayar ve kağıt-kalem formları arasında manidar bir fark bulunmuştur. Tüm çalışmalardan elde edilen ortak etki büyüklüğünün -.09 olduğu görülmüştür. Bu durum, kağıt-kalem ve bilgisayar ortamında uygulanan testlerden elde edilen puanlar arasındaki farkın .09 standart sapmaya denk olduğunu belirtmektedir – ki bu oldukça küçük bir etki büyüklüğünü gösterir. Dolayısıyla meta-analiz sonuçlarına göre kağıt-kalem ve bilgisayar ortamında uygulanan testlerde gösterilen öğrenci performansları arasında manidar bir fark yoktur. Bu durumda, bilgisayarlı testlere geçilmesinde bir sakınca görülmemekle birlikte, bilgisayarlı test uygulamaları geliştirilirken yapılacak deneme uygulamaları ile öğrenci performansının formlara göre farklılık gösterip göstermediğinin araştırılması gerekmektedir.

**Anahtar Sözcükler:** Bilgisayar ortamında uygulanan test, kağıt-kalem formu test, meta-analiz

**ABSTRACT:** Many studies examining on the comparison of computer-based tests and paper-and-pencil tests are available in the literature. These studies are applied to various fields, variables and workgroups. So it is important to compile and review these studies by using meta-analysis for researchers and authorities. In this paper, a survey of the studies on comparison of student performances in paper-based and computer-based tests is presented. By considering these studies, a meta-analysis on the comparison of students' achievements for paper-based and computer-based tests is provided. To determine if the studies included into meta-analysis have bias, Funnel Graph and Begg-Mazumdar correlation test are utilized for sensitivity analysis. Influence quantities

\* Arş. Gör., Osmangazi Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, ecaybek@gmail.com

\*\* Öğrt. Gör., Recep Tayyip Erdoğan Üniversitesi Eğitim Fakültesi, Eğitim Bilimleri Bölümü, bahar.0106@hotmail.com

\*\*\* Arş. Gör., Düzce Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü, merveeris87@gmail.com

\*\*\*\* Arş. Gör., Cumhuriyet Üniversitesi, Eğitim Fakültesi Eğitim Bilimleri Bölümü, asalihsimsek@gmail.com

\*\*\*\*\* Arş. Gör., Dumlupınar Üniversitesi Eğitim Fakültesi İlköğretim Bölümü, mustafakose43@gmail.com

<sup>1</sup> Bu çalışma, 29-31 Mayıs 2014 tarihleri arasında düzenlenen XIII. Ulusal Sınıf Öğretmenliği Sempozyumu'nda sözlü bildiri olarak sunulmuştur.

pertaining the studies are analyzed by using Comprehensive Meta-Analysis software. Results of heterogeneity test are analyzed to determine the best effect model and then it is seen that effect quantities shows normal distribution. Based on that, Fixed Effect Model is selected. Hedge's g coefficient is used to calculate the common effect. There are two basic restrictions of this study. First, due to the lack of time, it could not be possible to reach all studies available for the comparison of paper-based and computer-based tests. Second, results of the meta-analysis are not compared with respect to the topic fields. It could be better to compare the results for topic fields as future prospect. According to 35 findings related to the 9 different studies included to the meta-analysis, no reasonable difference is observed for paper-based and computer-based tests for 29 comparisons, and some reasonable differences are observed 6 ones. Common influence quantity obtained from all the studies is found as -0.09. This necessarily means that the differences between scores obtained from paper-based and computer-based tests corresponds to 0.09 standard deviation, which is not a considerable influence quantity. In other words, according to the meta-analysis results, there is no substantial difference for the students' performance in paper-based and computer-based tests. As a result, even though it is not inconvenient to use paper-based tests, it should be determined while preparing a computer-based test whether student achievement is affected by applied test form via some trial applications.

**Keywords:** Computer-based test, paper-based test, meta-analysis.

## GİRİŞ

Günümüzde bilgisayarlar neredeyse öğretimin bütün ilkelerini etkileyerek eğitimin bir parçası haline gelmeye başlamıştır. Hesaplama kaynaklarının kapasitesi ve imkânlarının artması eğitimde ölçme konusunda devrime yol açmıştır. Bilgisayar kullanarak yapılan testler; bir değerlendirme aracı olup; tek olarak, ağ ortamında veya internete bağlanarak bilgisayar tarafından yönetilirler. Bilgisayar ortamında uygulanan testler, 1970'lerin başından itibaren kullanılmaya başlamıştır (Drasgow, 2002). Ancak bu uygulamalar, kısıtlı bilgisayar kapasitesi ve yüksek maliyetler nedeniyle oldukça sınırlıdır. Bilgisayar ortamında uygulanan testler, yeni teknolojilerin gelişimiyle beraber lisans, sertifika, giriş ve psikolojik testler gibi geniş ölçekli test programlarında kullanılmıştır. Örneğin; Educational Testing Service (ETS)'in hazırladığı uluslararası geçerliğe sahip Graduate Record Examinations (GRE) birkaç yıldır bilgisayar ortamında olarak yürütülmektedir. TOEFL 1998'den itibaren bilgisayar ortamında uygulanmaktadır. Zaman içinde bilgisayar kullanılarak yapılan testlerin kalem-kağıt testlerinin yerini alması beklenmektedir. Örneğin; GRE testi artık kağıt-kalem formunda yapılmamaktadır.

Bilgisayar kullanılarak yapılan testlerin birçok avantaj ve dezavantajları bulunmaktadır. Bunlar şu şekilde özetlenebilir;

Öğrencileri sağladığı avantajlar,

1. Anında verilen dönüt ile öğrenme olanağını arttırması (Natal, 1998),
2. Yetersiz öğrencilere destek sağlaması, zamanın daha etkin kullanılması,
3. Yedi gün yirmi dört saat sınavı uygulayabilme imkanı bulunmasıdır (Luecht, 2001).

Eğitmenlere sağladığı avantajlar,

1. Kayıtların tutulması,
2. Test sorularının tekrar kullanılabilmesi,
3. Sınavların kolay okunması,
4. Çoklu ortamın (video, ses, resim...) eklenebilmesidir (Luecht, 2001; Natal, 1998).

Eğitmenler açısından bilgisayar ortamlarında uygulanan sınavların avantajları;

1. İstatistiksel dönüt alınabilmesi,
2. Ölçme hatalarını azaltması,
3. Geniş soru havuzlarının oluşturulabilmesi (Natal, 1998),
4. Soruların hemen güncellenebilmesi (Sampson, 2000),
5. Veri analizinin yapılmasıyla güvenilirliğin artması ve muhtemelen daha geçerli bir değerlendirme yapılabilmesidir (Helgeson & Kumar, 1993).

Öğrenciye dezavantajları;

1. Donanım veya erişimden kaynaklı süre problemi,
2. Öğrencinin bilgisayar erişimini gerektirmesi,
3. Bilgisayar becerileri gerektirmesi,
4. Öğrencilerin temel test stratejilerinden yoksun kalabilmesi (Natal, 1998),

5. Sistem hatasının meydana gelebilmesi,
6. Öğrenciye ek maliyet getirmesidir (Bugbee & Bernt, 1990).  
Eğitmenler için dezavantajı ise (Natal, 1998),
1. Güvenlik ve bilgisayar donanım ve yazılım sorunlarının meydana gelebilmesi,
2. Doğru dönüt ve puanlamanın programlama açısından sağlanabilmesidir.

Çalışmalar, test formatının sonuçları etkilemediğini ortaya koymuştur. Bilgisayar kullanarak yapılan testler, uygulama zamanı esnekliği, anında sonuçların ilan edilmesi ve güvenlik konularında geleneksel kağıt-kalem testlerine göre üstündür (Vrabel, 2004). Eğitimsel ve psikolojik ölçmelerde bilgisayarların kullanımı donanım maliyetlerinin fazla olması ve uygulama platformlarının değişmesine rağmen son otuz yılda sürekli artış göstermiştir. Bilgisayarların değerlendirmede oynadığı roller, birçok kapıyı açarken, beraberinde de birçok meseleyi getirmektedir. Bu meselelerin başında, geçerlik, cinsiyet farklılıkları, öğretimsel dağıtım, kullanıcı arabiriminin türü ve öğrenci başarısı konularıdır (Kumar, 1996).

Alanyazın incelendiğinde; bir çok çalışmanın geleneksel kağıt-kalem testleri ile bilgisayar kullanılarak uygulanan testleri karşılaştırdıkları görülmüştür. Örneğin; Kim ve Huynh (2007) tarafından yürütülen bir çalışmada, geniş ölçekli olarak eyalet çapında gerçekleştirilen ders sonu (EOC) cebir ve biyoloji sınavlarının kağıt ve kalem testleriyle bilgisayar kullanarak gerçekleştirilen uygulamalarında öğrenci puanlarını kıyaslamışlardır. Kalem-kağıt ve bilgisayar kullanılarak yapılan testler karşılaştırma yapılırken; ölçek puanları, madde parametre kestirimleri, karakteristik test eğrileri, test-bilgi fonksiyonu, baskın içerik düzeyine Rasch beceri kestirimleri ve yapının eşdeğerliğine göre incelenmiştir. Genel olarak, iki madde düzeyinde de madde düzeyinde, alt test düzeyinde ve bütün test düzeyinde ölçülen yapının bozulmadığına ilişkin bulgular elde edilmiştir.

Poggio, Glasnapp, Yang, ve Poggio (2005) tarafından yürütülen sözde-kontrollü ampirik bir çalışmada, Amerika Birleşik Devletleri'nde bir eyalette yürütülen geniş ölçekli değerlendirme programında öğrencilerin matematik başarısını ölçen bir testin bilgisayar ve kağıt kalem uygulamalarının öğrenci test puanlarına etkisi incelenmiştir. 7. sınıf öğrencileri bu çalışmanın evrenini oluşturmuştur. Gönüllülük esasına bağlı olarak 644 öğrenci önce bilgisayar daha sonra kağıt ve kalem test uygulamalarına katılmıştır. Tüm test göz önüne alındığında; bilgisayarda ve kağıt kalemle uygulanan testte öğrenci başarıları arasındaki fark manidar değildir. 204 maddenin dokuzunda fark bulunmuştur.

Özalp-Yaman ve Çağıltay (2010) tarafından yürütülen bir çalışmada, mühendislik eğitiminde kağıt ve kalem testleriyle bilgisayarda uygulanan testler karşılaştırılmıştır. Çalışmada, 209 mühendislik 1. Sınıf öğrencisinin kimya dersinde iki test formatındaki performansları incelenmiştir. Sonuçlara göre; İki test uygulamasında öğrenci başarılarına göre manidar bir fark bulunamamıştır.

Chan-Pensley (1999) tarafından yürütülen bir çalışmada, alkolikliği tespit etmek için kullanılan alkol-kullanım bozukluklarını tanıma testi (AUDIT) kullanılmıştır. Bir alkolden arınma etkinliğinde, hem bilgisayar hem de orijinal olan kağıt kalem uygulaması 110 gönüllü katılımcıya uygulanarak sonuçlar karşılaştırılmıştır. Çalışma için yaratılan bilgisayar versiyonunun kağıt kalem versiyonu kadar kabul edilebilir sonuçlar verdiği görülmüştür.

Boo ve Vispoel (2012) tarafından yürütülen bir çalışmada, Eğitimsel Gelişim Iowa Testleri'nin bilgisayar ve kağıt kalem versiyonlarında alınan puanlar değerlendirilmiştir. Bulgular, ortalama ve standart sapma, iç tutarlılık, ölçüt ve yapı geçerliğine göre incelenmiştir. Genel olarak sınava katılanlar bilgisayar uygulamalarına katılmayı tercih etmişler ve bu testlerin çalışma özelliklerini değerlendirmişlerdir. En olumlu tutum özellikleri, okuma beceri testleri ve kaydırılabilir okuma parçalarında belirlenmiştir.

Bayazit (2007) tarafından yürütülen bir çalışmada, çevrimiçi sınavlar ile kâğıt-kalem sınavları arasındaki başarı ve harcanan süre farklılığını incelenmiştir. Bir çevrimiçi sınav aracı geliştirilmiştir ve uygulanmıştır. Bu araç, çevrimiçi olarak veritabanı ile çalışan ve çoktan seçmeli (5 seçenekli, 4 seçenekli), doğru-yanlış, eşleştirme, boşluk doldurma, çok cevaplı, kısa cevaplı, uzun cevaplı soruların

eklenebilmesine, sınavların oluşturulmasına ve kâğıt-kalem sınavı formatına dönüştürülmesine olanak sağlayan bir araçtır. Başarım sınavı Hacettepe Üniversitesi Bilgisayar ve Öğretim Teknolojileri Eğitimi 3. sınıf öğrencilerine hem çevrimiçi hem de kâğıt kalem olarak uygulanmıştır. Araştırma kapsamında geliştirilen çevrimiçi sınav aracı, kullanım kolaylığı, amaca uygunluğu ve tasarımı açısından öğretim elemanlarınca değerlendirilmiştir. Öğrencilerin çevrimiçi sınav aracı ve çevrimiçi sınav ortamlarıyla ilgili görüşlerini belirlemek ve öğretim elemanlarının sınav aracı hakkındaki görüşlerini belirlemek amacıyla anketler geliştirilmiştir. Öğrencilerin kâğıt-kalem sınavından elde ettiği puanlar ile çevrimiçi sınavdan elde ettiği puanlar arasında anlamlı bir fark olmadığını göstermiştir. Ancak öğrencilerin çevrimiçi sınavda harcadıkları süreler, kâğıt-kalem sınavına göre daha fazladır. Öğrenciler, çevrimiçi sınav aracının kullanımı kolay bulmuşlardır ve çevrimiçi sınav ortamlarında kendilerini rahat hissettiklerini belirtmişlerdir. Ancak çevrimiçi sınav ortamlarıyla ilgili olarak seslerden rahatsız olmak, yorulmak ve sorulara odaklanamamak gibi şikâyetleri olmuştur.

Vispoel, Boo ve Bleiler (2001) tarafından yürütülen bir çalışmada, eğitim ve psikoloji araştırmalarında yaygın bir şekilde kullanılan Rosenberg Benlik Saygısı Ölçeği'nin (RBSÖ) bilgisayar ve kâğıt kalem versiyonlarının özellikleri değerlendirilmiştir. Sonuçlara göre; bilgisayar veya kâğıt kalem testi olması RBSÖ'nün psikometrik özelliklerini (puan büyüklüğü, değişkenlik ve faktör yapısı) çok az etkilemektedir. Bilgisayar versiyonu, daha uzun sürmekte ve katılımcılar tarafından daha çok tercih edilmektedir.

Pomplun, Frey ve Becker (2002) tarafından yürütülen bir çalışmada, bir okuma yerleştirme testinin bilgisayar ve kâğıt-kalem versiyonlarında puanların denkliği incelenmiştir. Bulgulara göre; iki bilgisayar versiyonunda da kâğıt kalem versiyonuna göre kelime-dağarcığı testi puanları daha yüksek bulunmuştur. Bir bilgisayar versiyonunda, hem kavrama testinde ve hem de genel testte daha yüksek puanlar elde edilmiştir. Bilgisayar versiyonundan elde edilen puanlar ders yerleştirme açısından kâğıt kalem versiyonuyla benzer yordama gücü göstermiştir.

Aybek (2012) tarafından yürütülen bir çalışmada, bir genel yetenek testi kâğıt-kalem ve bilgisayar ortamında uygulanarak, testin psikometrik özelliklerinin ve bilgisayara yönelik tutum, bilgisayar aşinalığı ve cinsiyete göre öğrenci performanslarının karşılaştırılması amaçlanmıştır. Çalışmanın araştırma grubunu Eskişehir'de ilköğretim okullarından seçilen; 73 erkek, 63 kız toplam 136 6. sınıf öğrencisi oluşturmuştur. Araştırma grubuna sözel ve sayısal akıl yürütme alt testlerinden oluşan, toplam 60 maddelik genel yetenek testi kâğıt-kalem ve bilgisayar ortamında uygulanmıştır. Araştırma grubuna aynı zamanda bilgisayara yönelik tutum ölçeği ve bilgisayar aşinalığı ölçekleri de verilmiştir. Araştırmadan elde edilen bulgular; testin kâğıt kalem ve bilgisayar ortamında uygulanan formlarının büyük ölçüde eşdeğer olduğunu; bilgisayar aşinalığı ve cinsiyete göre, testin bilgisayar ortamında uygulanan formundan alınan puanların farklılaşmadığını; buna karşın, bilgisayara yönelik tutuma göre bilgisayar ortamında uygulanan formdan alınan puanların farklılaştığını göstermiştir.

Bilgisayar kullanılarak uygulanan testlerle geleneksel kâğıt-kalem testlerini karşılaştıran bir çok araştırma bulunmaktadır. Bu araştırmalar çok çeşitli alanlara, farklı değişkenlere, farklı çalışma gruplarına yayılmıştır. Bu çalışmaların sonuçlarının meta-analiz yardımıyla toparlanıp derlenmesi hem diğer araştırmacılar hem de karar alıcı yetkililer açısından önemlidir.

## YÖNTEM

Araştırmada, kâğıt-kalem formu ve bilgisayar ortamında uygulanan testlerde gösterilen öğrenci performanslarını karşılaştıran çalışmalar taranmıştır. Bu çalışmalar üzerinde kâğıt-kalem formu ve bilgisayar ortamında uygulanan testlerin öğrenci performansı bakımından karşılaştırılması ile ilgili meta analiz çalışması yapılmıştır. Analiz sonucunda testlerden performansı belirlemeye ilişkin bilgi edinilmesi amaçlandığı için betimsel bir araştırma olup, "tarama" modelindedir.

Çalışma sırasında, araştırmanın amacına uygun olarak, meta-analizin Durlak (1995) tarafından tanımlanmış olan altı temel adımı dikkate alınmıştır. Bu adımlar aşağıda sıralanmıştır;

1. Araştırma sorusunu açık ve anlaşılır bir şekilde ortaya koyma

2. Literatür tarama
3. Çalışmayı kodlama
4. Etki indeksi belirleme
5. İstatistiksel analizler
6. Sonuçlar ve yorumlar

Bu araştırmanın çalışma grubunu, kağıt kalem formu ve bilgisayar ortamında uygulanan testlerin öğrenci performansını belirlemeye yönelik yapılan çalışmalar oluşturmaktadır. Ulaşılan 12 çalışma;

1. Bilgisayar ortamında uygulanan testler
2. Kağıt-kalem testleri
3. Performans karşılaştırma

ölçütlerine göre değerlendirilmiştir. Ölçütlere uygun olmayan 3 çalışma meta analize dahil edilmemiş, ölçütlere uygun olan 9 çalışma meta analize dahil edilmek üzere belirlenmiştir. Ölçütlere uygun olan çalışmalar çalışma grubunu oluşturmaktadır. Analize dahil edilen çalışmalar Tablo 1’de verilmiştir.

**Tablo 1: Meta-Analize Dahil Edilen Çalışmalar**

Çalışma Kodu	Çalışma Adı
Chan-Pensley (1999)	Alcohol-Use Disorders Identification Test: A Comparison Between Paper And Pencil And Computerized Versions
Kim & Hyunh (2007)	Comparability of Computer and Paper and-Pencil Versions of Algebra and Biology Assessments
Vispoel (2001)	Computerized And Paper-And-Pencil Versions Of The Rosenberg Self-Esteem Scale: A Comparison Of Psychometric Features And Respondent Preferences
Poggio ve diğ. (2005)	A Comparative Evaluation of Score Results from Computerized and Paper & Pencil Mathematics Testing in a Large Scale State Assessment Program
Pomplun ve diğ. (2002)	The Score Equivalence Of Paper-And-Pencil And Computerized Versions Of A Speeded Test Of Reading Comprehension
Aybek (2012)	Kağıt - Kalem Formu Ve Bilgisayar Ortamında Uygulanan Genel Yetenek Testinin Psikometrik Özelliklerinin Karşılaştırılması
Boo ve Vispoel (2012)	Computer versus Paper-and-Pencil Assessment of Educational Development: A Comparison of Psychometric Features and Examinee Preferences I
Yaman & Çağıltay (2010)	Paper-Based versus Computer-Based Testing in Engineering Education
Bayazıt ( 2007)	Çevrimiçi Sınavlar Ve Kâğıt-Kalem Sınavları Arasındaki Sınav Süresi Ve Öğrenci Başarım Farklılıkları

Meta analiz yönteminde etki büyüklükleri birbirinden farklı ögelerin analize dahil edilmesi gerektiği için, buradaki analiz de etki büyüklükleri birbirinden farklı çalışmaları içermektedir. Önemli olan bu farklılıkların makul düzeyde olmasıdır. Çalışmaların istatistiksel olarak birleştirilebilmesi için normal dağılıma uygun olup olmadığının belirlenmesi gerekmektedir.

Bir meta analiz çalışmasında kullanılacak olan araştırmaların seçimi ayrıntılı kaynak taraması gerektirmektedir. Geçerli ve güvenilir sonuçlara ulaşabilmek için konu ile ilgili yurt içi ve yurt dışında yapılan çalışmaların tümüne ulaşmaya çalışılmalıdır. Ardından araştırmacılar tarafından belirlenen meta analize dahil edilme ölçütlerine uygun çalışmaların analize dahil edilmesi gerekmektedir. Bu araştırmada çalışmaların çoğunluğu yurtdışında yapılmış olup internet arama motorları, ULAKBİM veri tabanı ve üniversite kütüphanelerinin veri tabanlarından yararlanılmıştır.

Meta analize dahil edilen çalışmaların yanlı olup olmadığını belirlemek için ise duyarlılık analizi olarak Funnel grafiği ve Begg-Mazumdar sıra korelasyonları testinden yararlanılmıştır. İncelenen çalışmalara ait etki büyüklükleri Comprehensive Meta Analysis yazılımı ile incelenmiştir. Hangi etki modelinin benimseneceğine karar vermek amacıyla heterojenlik testi sonuçları incelenmiş ve çalışmalara ait etki büyüklüklerinin normal dağılım gösterdiği görülmüştür ( $Q_{34} = 43.08$ ;  $p > .05$ ). Bu bağlamda sabit etki modelinin seçilmesinde bir sakınca görülmemiştir. Ortak etki büyüklüğü hesaplamasında Hedge'in g katsayısının kullanılması uygun görülmüştür. Hedge'in g katsayısı, grup büyüklükleri eşit olmadığında kullanılmaktadır ve her grubun standart sapmasını, grup büyüklüğüne göre ağırlıklandırmaktadır.

## BULGULAR

Meta-analize dahil edilen 9 farklı çalışmaya ait toplam 35 bulguya göre, 29 karşılaştırmada bilgisayar ve kağıt-kalem formları arasında manidar bir fark bulunmazken, 6 çalışmada bilgisayar ve kağıt-kalem formları arasında manidar bir fark bulunmuştur.

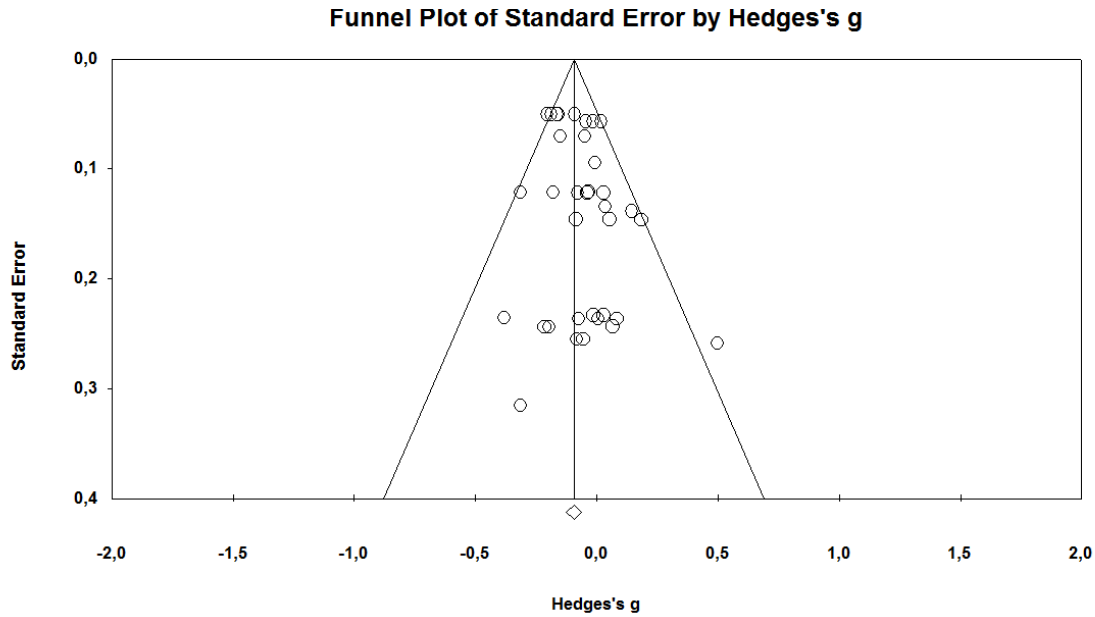
Meta-analize dahil edilen çalışmaların etki büyüklüklerine ilişkin bilgiler Tablo 2'de yer almaktadır. Buna göre, tüm çalışmalardan elde edilen ortak etki büyüklüğünün  $-.09$  olduğu görülmektedir. Bu durum, kağıt-kalem ve bilgisayar ortamında uygulanan testlerden elde edilen puanlar arasındaki farkın  $.09$  standart sapmaya denk olduğu görülmektedir –ki bu oldukça küçük bir etki büyüklüğüne işaret etmektedir. Dolayısıyla meta-analiz sonuçlarına göre kağıt-kalem ve bilgisayar ortamında uygulanan testlerde gösterilen öğrenci performansları arasında manidar bir fark olmadığı görülmektedir.

**Tablo 2: Meta-Analiz Sonuçları**

Çalışma Adı	Hedge g	Standart		Alt Sınır	Üst Sınır	Z	p
		Hata	Varyans				
Chan-Pensley (1999)	0,03	0,13	0,02	-0,23	0,30	0,26	0,80
Vispoel (2001)	-0,01	0,09	0,01	-0,19	0,18	-0,06	0,95
Kim & Hyunh (2007)	-0,05	0,07	0,00	-0,19	0,09	-0,68	0,49
Kim & Hyunh (2007)	-0,15	0,07	0,00	-0,29	-0,01	-2,12	0,03
Kim & Hyunh (2007)	-0,19	0,05	0,00	-0,29	-0,09	-3,71	0,00
Kim & Hyunh (2007)	-0,16	0,05	0,00	-0,26	-0,06	-3,18	0,00
Kim & Hyunh (2007)	-0,20	0,05	0,00	-0,30	-0,10	-4,02	0,00
Kim & Hyunh (2007)	-0,16	0,05	0,00	-0,26	-0,06	-3,19	0,00
Kim & Hyunh (2007)	-0,09	0,05	0,00	-0,19	0,01	-1,82	0,07
Poggio ve diğ. (2005)	-0,05	0,06	0,00	-0,16	0,07	-0,79	0,43
Poggio ve diğ. (2005)	-0,02	0,06	0,00	-0,13	0,10	-0,27	0,79

Poggio ve diğ. (2005)	0,02	0,06	0,00	-0,09	0,13	0,33	0,74
Pomplun ve diğ. (2002)	-0,08	0,15	0,02	-0,37	0,20	-0,58	0,56
Pomplun ve diğ. (2002)	0,18	0,15	0,02	-0,10	0,47	1,26	0,21
Pomplun ve diğ. (2002)	0,05	0,15	0,02	-0,23	0,34	0,37	0,71
Aybek (2012)	-0,03	0,12	0,01	-0,27	0,20	-0,28	0,78
Aybek (2012)	-0,31	0,12	0,01	-0,55	-0,07	-2,57	0,01
Aybek (2012)	-0,18	0,12	0,01	-0,42	0,06	-1,49	0,14
Boo ve Vispoel (2012)	-0,01	0,23	0,05	-0,47	0,44	-0,05	0,96
Boo ve Vispoel (2012)	-0,07	0,24	0,06	-0,54	0,39	-0,31	0,76
Boo ve Vispoel (2012)	-0,06	0,25	0,06	-0,55	0,44	-0,22	0,83
Boo ve Vispoel (2012)	-0,19	0,24	0,06	-0,67	0,28	-0,80	0,43
Boo ve Vispoel (2012)	-0,08	0,12	0,01	-0,32	0,16	-0,64	0,52
Boo ve Vispoel (2012)	-0,38	0,24	0,06	-0,84	0,08	-1,62	0,11
Boo ve Vispoel (2012)	0,00	0,24	0,06	-0,46	0,47	0,02	0,98
Boo ve Vispoel (2012)	0,50	0,26	0,07	-0,01	1,01	1,92	0,05
Boo ve Vispoel (2012)	-0,21	0,24	0,06	-0,69	0,26	-0,88	0,38
Boo ve Vispoel (2012)	-0,04	0,12	0,01	-0,28	0,20	-0,32	0,75
Boo ve Vispoel (2012)	0,03	0,23	0,05	-0,43	0,49	0,13	0,90
Boo ve Vispoel (2012)	0,08	0,24	0,06	-0,38	0,55	0,35	0,72
Boo ve Vispoel (2012)	-0,08	0,25	0,07	-0,58	0,42	-0,32	0,75
Boo ve Vispoel (2012)	0,07	0,24	0,06	-0,41	0,54	0,28	0,78
Boo ve Vispoel (2012)	0,03	0,12	0,01	-0,21	0,27	0,24	0,81
Yaman & Çağıltay (2010)	0,15	0,14	0,02	-0,13	0,42	1,05	0,29
Bayazıt ( 2007)	-0,31	0,32	0,10	-0,93	0,30	-0,99	0,32
<b>Sabit Etki Modeli</b>	<b>-0,09</b>	<b>0,02</b>	<b>0,00</b>	<b>-0,12</b>	<b>-0,06</b>	<b>-6,17</b>	<b>0,00</b>

Meta-analize dahil edilen çalışmaların yanlı olup olmadığını belirlemek için Funnel grafiği (Şekil 1) incelenmiştir.



Funnel grafiğinin asimetrik olması, çalışma örnekleminin yanlı olduğu anlamına gelmektedir. Ancak elde edilen grafiğin asimetrik olmadığı görülmektedir. Çalışma yanlılığını belirlemenin diğer bir yolu ise Begg ve Mazumdar sıra korelasyonları testidir. Buna göre, Kendall'ın tau b katsayısı hesaplanır ve bu katsayının manidar olması durumunda, çalışmada yanlılık olduğu anlamına gelmektedir. Bu bağlamda; meta-analize dahil edilen çalışma örnekleminin yanlı olmadığı anlaşılmaktadır (tau b = .06; p > .05).

## SONUÇ VE ÖNERİLER

Yapılan meta-analiz sonucuna göre, kağıt-kalem formu ve bilgisayar ortamında uygulanan testler bir bütün olarak ele alındığında, öğrenci başarısında farklılığa yol açtığı (Hedge g = -.09; p < .05) bulunmuştur ve bu bulgu Pomplun, Frey ve Becker (2002) tarafından yürütülen deneysel çalışmada elde edilen bulgularla örtüşür niteliktedir. Ancak hesaplanan ortak etki büyüklüğünün oldukça küçük olması nedeniyle bu farkın göz ardı edilebileceği söylenebilir ki bu da Poggio, Glasnapp, Yang, ve Poggio (2005) tarafından yürütülen sözde-kontrollü ampirik çalışmanın ve Özalp-Yaman ve Çağıltay (2010) tarafından yürütülen çalışmanın bulguları ile örtüşür niteliktedir.

Bu araştırmanın zamansal sınırlılıkları nedeniyle, kağıt-kalem formunda ve bilgisayar ortamında uygulanan testlere ait karşılaştırmaların yapıldığı çalışmaların tümüne ulaşmak mümkün olmamıştır. Yanlılık testi sonuçları her ne kadar araştırmaya dahil edilen çalışmaların yanlı olmadığını ortaya koysa da, ulaşılamayan çalışmaların da dikkate alındığı bir meta-analiz araştırmasının farklı bulgular ortaya koyabileceği unutulmamalıdır. Yine bu çalışmanın bir sınırlılığı olarak, meta-analiz sonuçları konu alanlarına göre karşılaştırılamamıştır. Gelecek araştırmalarda konu alanlarına göre meta-analiz sonuçlarının karşılaştırılması faydalı olacaktır.

Meta-analiz sonuçlarının gösterdiği üzere, kağıt-kalem formu ve bilgisayar ortamında uygulanan testlerde gösterilen öğrenci performansları arasında manidar bir fark bulunmamaktadır. Bu durumda, bilgisayarlı testlere geçilmesinde bir sakınca görülmemekle birlikte, bilgisayarlı test uygulamaları geliştirilirken yapılacak deneme uygulamaları ile öğrenci performansının formlara göre farklılık gösterip göstermediğinin araştırılması gerekmektedir.

## KAYNAKÇA

Aybek, E. C. (2012). *Kağıt-Kalem Formu ve Bilgisayar Ortamında Uygulanan Genel Yetenek Testinin Psikometrik Özelliklerinin Karşılaştırılması*. (Yayınlanmamış yüksek lisans tezi). Ankara Üniversitesi, Ankara.

Bayazıt, A. (2007). *Çevrimiçi Sınavlar Ve Kağıt-Kalem Sınavları Arasındaki Sınav Süresi ve Öğrenci Başarım Farklılıkları*. (Yayınlanmamış yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.

Boo, J., & Vispoel, W. (2012). Computer versus paper-and-pencil assessment of educational development: a comparison of psychometric features and examinee preferences. *Psychological Reports, 111*(2), 443–460. doi:10.2466/10.03.11.PR0.111.5.443-460.

Bugbee Jr., A. C., & Bernt, F. M. (1990). Testing by computer: findings in six years of use 1982-1988. *Journal of Research on Computing in Education, 23*(1), 87-101.

Chan-Pensley, E. (1999). Alcohol-use disorders identification test: a comparison between paper and pencil and computerized versions. *Alcohol and Alcoholism (Oxford, Oxfordshire), 34*(6), 882–885. 04.04.2013 tarihinde <http://www.ncbi.nlm.nih.gov/pubmed/10659724> adresinden alınmıştır.

Drasgow, F (2002). The work ahead: a psychometric infrastructure for computerized adaptive tests. In C.N. Mills, M.T. Potenza, J.J. Fremer, & W.C. Ward (Eds.), *Computer-based testing: Building the foundation for future assessments* (pp. 67–88). Hillsdale, NJ: Lawrence Erlbaum.



Durlak, J.A. (1995). Understanding meta-analysis. In L.G. Grimm, & P.R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 319-352). Washington, DC: American Psychological Association.

Helgeson, S. L., & Kumar, D. D. (1993). A review of educational technology in science assessment. *National Center for Science Teaching and Learning*, Columbus, Monograph Series Number 7, OH., (ERIC Document Reproduction Service No. ED366507).

Kim, D., & Huynh, H. (2007). Comparability of computer and paper-and-pencil versions of algebra and biology assessments. *The Journal of Technology, Learning and Assessment*, 6(4), 1–29. 08.05.2013 tarihinde <http://napoleon.bc.edu/ojs/index.php/jtla/article/view/1634> adresinden alınmıştır.

Kumar, D. (1996). *Computers and Assessment in Science Education*. ERIC Clearinghouse for Science, Mathematics and Environmental Education, Columbus, OH., (ERIC Document Reproduction Service No. ED395770).

Luecht, R. M. (2001). *Challenges of Web-Based Assessment*. (ERIC Document Reproduction Service No. ED458294).

Natal, D. (1998). *OnLine assessment: what, why, how*. *Imagen Multimedia Corp.*, Lompoc, CA., (ERIC Document Reproduction Service No. ED419552).

Özalp-Yaman, Ş., & Çağiltay, N. E. (2010). Paper-based versus computer-based testing in engineering education. *IEEE EDUCON Education Engineering* (pp. 1631–1637). 05.05.2013 tarihinde [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5492397](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5492397) adresinden alınmıştır.

Poggio, J., & Glasnapp, D. (2005). A comparative evaluation of score results from computerized and paper & pencil mathematics testing in a large scale state assessment program. *The Journal of Technology, Learning and Assessment*, 3(6), 1–30.

Pomplun, M., Frey, S., & Becker, D. F. (2002). The Score Equivalence of Paper-and-Pencil and Computerized Versions of a Speeded Test of Reading Comprehension. *Educational and Psychological Measurement*, 62(2), 337–354. doi:10.1177/0013164402062002009

Sampson, J.P. (2000). Using the internet to enhance testing in counseling. *Journal of Counseling and Development*, 78(3), 348-356.

Vispoel, W. P., Boo, J., & Bleiler, T. (2001). Computerized and Paper-and-Pencil Versions of the Rosenberg Self-Esteem Scale: A Comparison of Psychometric Features and Respondent Preferences. *Educational and Psychological Measurement*, 61(3), 461–474. doi:10.1177/00131640121971329

Vrabel M. (2004). Computerized versus paper-and-pencil testing methods for a nursing certification examination: a review of the literature. *Computers, Informatics, Nursing*, 22(2), 94-98.