

Laboratuvar mı, Gerçek Dünya mı? Türkçe Web'de Lighthouse ve CrUX Ölçümlerinin Uyumu Üzerine Bir Yöntem-Karşılaştırma Çalışması

Lab or Field? A Method-Comparison Study of Lighthouse and CrUX Measurements on the Turkish Web

1st.com.tr Araştırma Birimi, 1st.com.tr, İstanbul, Türkiye

DOI: <https://doi.org/10.5281/zenodo.20732713> · **Yayın:** 10 Haziran 2026 · **Seri:** TR-WEB Çalışma Kâğıtları, 1 (2026) · **Lisans:** CC BY 4.0

Öz

Amaç: Laboratuvar (Lighthouse, kısıtlı/throttled) performans ölçümlerinin gerçek-kullanıcı saha (Chrome UX Report, CrUX p75) deneyimini ne ölçüde öngördüğünü belirlemek. **Yöntem:** Aynı PageSpeed Insights çağrısından laboratuvar (Lighthouse) ve saha (CrUX p75) verisi alınan 7.403 Türkçe site için En Büyük İçerikli Boyama (LCP), İlk İçerikli Boyama (FCP) ve Kümülatif Düzen Kayması (CLS) eşleştirilmiş; uyum Pearson ve Spearman korelasyonu, Bland-Altman analizi ve Cohen's kappa (Core Web Vitals 'iyi/geliştirilmeli/zayıf' sınıflaması) ile incelenmiştir. **Bulgular:** Laboratuvar, yükleme metriklerini sistematik ve büyük ölçüde kötümser raporlamıştır (LCP: lab medyan 10.876 ms, saha 2.247 ms). LCP'de lab-saha ilişkisi zayıftır ($p = 0,26$) ve sınıf uyumu şanstan ileri gitmemektedir ($\kappa \approx 0,00$); sitelerin %85,6'sı laboratuvar da gerçek dünyadan daha kötü bir sınıfa düşmüş, laboratuvarın 'iyi' oranı %3,1 iken sahada bu oran %59,6 olmuştur. FCP benzer biçimde zayıf uyum ($\kappa = 0,01$) verirken, yalnız CLS orta düzeyde uyum göstermiştir ($\kappa = 0,33$). Lighthouse genel skoru saha LCP'siyle yalnız zayıf korelasyon kurmuştur ($r = -0,25$). **Sonuç:** Türkçe web'de varsayılan laboratuvar Lighthouse yükleme metrikleri, gerçek-kullanıcı deneyiminin zayıf bir vekilidir; performans kararları saha verisine dayandırılmalı, laboratuvar ise teşhis/hata ayıklama amacıyla kullanılmalıdır.

ABSTRACT

Objective: To determine how well laboratory (Lighthouse, throttled) performance measurements predict real-user field (Chrome UX Report, CrUX p75) experience. **Method:** For 7,403 Turkish sites with laboratory (Lighthouse) and field (CrUX p75) values from the same PageSpeed Insights call, Largest Contentful Paint (LCP), First Contentful Paint (FCP) and Cumulative Layout Shift (CLS) were matched; agreement was assessed with Pearson and Spearman correlations, Bland-Altman analysis, and Cohen's kappa over the Core Web Vitals 'good/needs-improvement/poor' classification. **Results:** The laboratory systematically and substantially over-reported loading metrics (LCP: lab median 10,876 ms vs field 2,247 ms). The lab-field relationship for LCP was weak ($p = 0.26$) and classification agreement did not exceed chance ($\kappa \approx 0.00$); 85.6% of sites were classified worse in the lab than in the field, and while only 3.1% of sites were 'good' in the lab, 59.6% were 'good' in the field. FCP showed similarly weak agreement ($\kappa = 0.01$); only CLS reached moderate agreement ($\kappa = 0.33$). The overall Lighthouse score correlated only weakly with field LCP ($r = -0.25$). **Conclusion:** On the Turkish web, default laboratory Lighthouse loading metrics are a poor proxy for real-user experience; performance decisions should rest on field data, with the laboratory reserved for diagnosis.

Anahtar kelimeler: Core Web Vitals, Lighthouse, CrUX, laboratuvar-saha uyumu, Bland-Altman, Cohen's kappa, LCP, yöntem karşılaştırma

1. Giriş

Web performansı iki temel kaynaktan ölçülür: kontrollü, sentetik laboratuvar testleri (ör. Lighthouse) ve gerçek kullanıcılardan toplanan saha verisi (ör. Chrome UX Report, CrUX). Laboratuvar testi tek bir kısıtlı (throttled) koşumda tekrarlanabilir bir tanı sağlarken; saha verisi gerçek cihaz, ağ ve coğrafya dağılımının p75 (75. yüzdelik) özetini verir. Bu ikisi ilişkili ama özdeş olmayan büyüklüklerdir.

Uygulamada birçok ekip kararlarını yalnız laboratuvar skorlarına dayandırır; oysa laboratuvarın gerçek deneyimi ne ölçüde temsil ettiği ampirik bir sorudur. Bu çalışma şu soruyu yanıtlatır: Türkçe web'de laboratuvar Lighthouse ölçümleri, gerçek-kullanıcı saha deneyiminin geçerli bir vekili midir? Soruyu tek bir özelliği (Core Web Vitals) merkeze alarak, yöntem-karşılaştırma istatistikleriyle yanıtıyoruz.

2. İlgili Çalışmalar

Core Web Vitals (LCP, FCP, CLS, INP) kullanıcı deneyiminin yerleşik göstergeleridir; Google'ın rehberliği saha verisini 'gerçek' ölçüt, laboratuvarı ise tanı aracı olarak konumlar. İki ölçüm yönteminin uyumunu sinamak için tıbbi ölçüm literatüründen gelen Bland & Altman (1986) yaklaşımı (yanlılık + uyum sınırları) ve kategorik uyum için Cohen (1960) kappa katsayısı standarttır; kappa yorumunda Landis & Koch (1977) eşikleri kullanılır. Bu çalışmanın katkısı, anılan yöntem-karşılaştırma çerçevesini büyük bir Türkçe site popülasyonunda laboratuvar-saha CWV uyumuna uygulamasıdır.

3. Yöntem

Örneklem, 1st.com.tr analiz motorunca denetlenen ve hem laboratuvar hem CrUX saha verisine sahip 7.403 Türkçe siteden oluşur (saha verisi tüm denetlenen 13.775 sitenin %53,7'sinde mevcuttu). Hem laboratuvar hem saha değerleri tek bir PageSpeed Insights (PSI) API çağrısından (strategy=mobile) elde edilmiştir: laboratuvar değerleri PSI'nin koşturduğu Lighthouse'tan (lighthouseResult; mobil cihaz emülasyonu, ağ ve CPU kısıtlaması ile), saha değerleri ise aynı yanıtındaki CrUX alan verisinden (loadingExperience/originLoadingExperience; 28-günlük p75). CrUX verisi yalnızca başarılı bir PSI yanıtında bulunduğundan, eşleşen örneklemdaki her site hem laboratuvar hem saha sayısını aynı PSI çağrısından alır; dolayısıyla karşılaştırma, PageSpeed Insights'ın aynı origin için yan yana sunduğu laboratuvar ve saha değerleri üzerinedir. Her site için LCP, FCP (milisaniye) ve CLS (birimsiz) laboratuvar-saha çiftleri eşleştirilmiştir.

Önemli bir ayrım: laboratuvar tek bir sentetik koşumu, saha ise gerçek kullanıcı dağılımının p75'ini ölçer. Dolayısıyla karşılaştırma 'aynı şeyi mi ölçüyorlar' değil, 'laboratuvar sahanın kullanışlı bir vekili mi' sorusudur.

4. İstatistiksel Analiz

Üç tamamlayıcı yöntem kullanıldı. (1) İlişki: Pearson (p1-p99 winsorize, uç değer etkisini azaltmak için) ve dağılımdan bağımsız Spearman korelasyonu; her katsayı %95 güven aralığı (Fisher z) ve iki-yönlü p ile. (2) Uyum: Bland-Altman analizi; yanlılık (lab – saha ortalama farkı) ve %95 uyum sınırları (yanlılık $\pm 1,96 \cdot SD$); uç değerlere karşı medyan mutlak fark da raporlandı. (3) Sınıf uyumu: Core Web Vitals eşikleriyle (LCP 2.500/4.000 ms; FCP 1.800/3.000 ms; CLS 0,10/0,25) her ölçüm 'iyi/geliştirilmeli/zayıf' olarak sınıflanıp Cohen's kappa hesaplandı (yorum: Landis & Koch 1977). Büyük örneklemden anlamlılık neredeyse garanti olduğundan yorum, etki/uyum büyüklüğüne dayandırıldı.

5. Bulgular: Laboratuvar ve Saha Değerleri

Laboratuvar, yükleme metriklerini sahaya kıyasla çarpıcı biçimde yüksek (kötümser) raporlamıştır (Tablo 1). LCP'de laboratuvar medyanı saha medyanının yaklaşık beş katıdır; FCP'de fark daha küçük ama belirgindir. CLS'de iki kaynak birbirine en yakın değerleri vermiştir.

METRİK	LAB MEDYAN	SAHA MEDYAN	LAB ORT.	SAHA ORT.
LCP (ms)	10.876	2.247	14.769	2.515
FCP (ms)	3.785	1.649	4.466	1.941
CLS	0,008	0,010	0,130	0,096

Tablo 1. Eşleşen sitelerde laboratuvar ve saha merkezî eğilim değerleri ($n = 7.403$). LCP/FCP ms, CLS birimsiz.

6. Bulgular: Korelasyon

Laboratuvar ve saha değerleri arasındaki ilişki, yükleme metriklerinde zayıftır: LCP için Spearman $\rho = 0,26$, FCP için $\rho = 0,17$. Yalnız düzen kararlılığı (CLS) orta düzeyde bir ilişki göstermiştir ($\rho = 0,41$). Lighthouse genel performans skoru, saha LCP ve INP'siyle beklenen yönde fakat zayıf korelasyon kurmuştur ($r \approx -0,25$): yüksek laboratuvar skoru, gerçek kullanıcıda daha iyi değerlerle yalnızca gevşek biçimde ilişkilidir (Tablo 2).

İlişki	PEARSON R	%95 GA	SPEARMAN P	P
LCP (lab ↔ saha)	0,21	[0,19; 0,23]	0,26	<0,0001
FCP (lab ↔ saha)	0,15	[0,13; 0,17]	0,17	<0,0001
CLS (lab ↔ saha)	0,43	[0,41; 0,45]	0,41	<0,0001
LH skoru ↔ saha LCP	-0,25	[-0,27; -0,23]	-0,28	<0,0001
LH skoru ↔ saha INP	-0,25	[-0,27; -0,22]	-0,27	<0,0001

Tablo 2. Laboratuvar-saha korelasyonları (Pearson winsorize; Spearman robust). GA: Fisher z %95 güven aralığı.

7. Bulgular: Bland-Altman Uyum

Bland-Altman analizi, yükleme metriklerinde geniş ve uygulanamaz bir uyumsuzluk ortaya koyar (Tablo 3). LCP'de laboratuvar, sahayı ortalama 12.254 ms aşmakta (yanlılık); %95 uyum sınırları -18,4 sn ile +42,9 sn arasında uzanmaktadır. Yani tek bir laboratuvar LCP değerinden gerçek saha LCP'si pratikte kestirilemez. Uç değerlere karşı sağlam olan medyan mutlak fark dahi LCP'de 8.384 ms'dir. CLS'de yanlılık küçüktür (+0,034) ama uyum sınırları yine geniştir.

METRİK	YANLILIK	%95 UYUM SINIRLARI	MEDYAN FARK
LCP (ms)	+12.254	[-18.377; +42.886]	8.384
FCP (ms)	+2.524	[-3.738; +8.786]	2.050
CLS	+0,034	[-0,46; +0,53]	0,02

Tablo 3. Bland-Altman uyum analizi (lab – saha). Yanlılık: ortalama fark; uyum sınırları: yanlılık $\pm 1,96 \cdot SD$.

8. Bulgular: Sınıf Uyumu (Cohen's Kappa)

Core Web Vitals 'iyi/geliştirilmeli/zayıf' sınıflamasında laboratuvar ile saha arasındaki uyum, yükleme metriklerinde şanstıan ileri gitmemektedir (Tablo 4). LCP için $\kappa = 0,00$: laboratuvar sınıflaması saha sınıflamasını rastlantıdan daha iyi öngörmez. En çarpıcı sonuç asimetridir: sitelerin %85,6'sı laboratuvar da gerçeğinden daha kötü bir sınıfa düşmüş; laboratuvar da yalnızca %3,1 site 'iyi' LCP'ye sahip görünürken, sahada bu oran %59,6'dır. FCP'de durum benzerdir ($\kappa = 0,01$). Yalnız CLS orta düzeyde uyum gösterir ($\kappa = 0,33$; siteler %72,8 oranında aynı sınıfta).

METRİK	K	LAB DAHA KÖTÜ %	AYNI SINIF %	LAB 'İYİ' %	SAHA 'İYİ' %
LCP	0,00	85,6	13,1	3,1	59,6
FCP	0,01	75,4	20,5	7,7	58,3
CLS	0,33	16,1	72,8	73,3	76,5

Tablo 4. CWV sınıf uyumu (Cohen's kappa). 'Lab daha kötü %': laboratuvarın sahadan daha kötü sınıfa koyduğu site oranı.

9. Tartışma

Bulgular tutarlı bir tablo çizer: Türkçe web'de varsayılan laboratuvar Lighthouse yükleme metrikleri, gerçek-kullanıcı deneyiminin zayıf bir vekilidir. Zayıf korelasyon (LCP $\rho = 0,26$), uygulanamaz Bland-Altman uyum sınırları ve şans düzeyinde kappa (LCP $\kappa = 0,00$) aynı sonuca işaret eder. Laboratuvarın sistematik kötümserliği, varsayılan mobil-kısıtlı (throttled) koşum, soğuk ön bellek ve tek-koşumun p75 ile özdeş olmaması gibi nedenlerle açıklanabilir.

Düzen kararlılığı (CLS) istisna oluşturur: laboratuvar ve saha orta düzeyde uyuşur ($\kappa = 0,33$), çünkü CLS büyük ölçüde ağ koşullarından bağımsız, belgenin yapısal bir özelliğidir. Pratik çıkarım nettir: performans kararları ve sıralamalar saha (CrUX) verisine dayandırılmalı; laboratuvar ise regresyonu yerelde teşhis etmek için kullanılmalıdır. Laboratuvar genel skoruyla sahayı ($r \approx -0,25$) sıralamak, sitelerin çoğunu yanlış konumlandırır.

10. Sınırlılıklar

(i) Laboratuvar tek bir sentetik koşumu, saha ise p75'i ölçer; bu büyüklük (estimand) farkı, gözlenen uyumsuzluğun bir bölümünü doğal olarak açıklar. Ancak uygulama açısından soru yine 'laboratuvar kullanışlı bir vekil mi' olduğundan bulgu geçerlidir. (ii) Saha verisi yalnızca yeterli trafiğe sahip %53,7'lik alt-kümede mevcuttur (seçilim yanlılığı). (iii) Laboratuvar varsayılan yapılandırmayla alınmıştır; farklı throttling ayarları farklı yanlılık verebilir. (iv) Çalışma gözlemsel ve kesitseldir. (v) Uç değerler nedeniyle birincil yorum, sağlam ölçütlere (Spearman, medyan fark, kappa) dayandırılmıştır.

11. Etik ve Veri Kullanımı

Tüm ölçümler kamuya açık sayfaların otomatik denetimine ve kamuya açık CrUX toplulaştırmalarına dayanır; kişisel veri toplanmaz ve sonuçlar toplulaştırılmıştır. Çalışma CC BY 4.0 ile açık erişimle yayımlanır; veri akademik amaçla talep üzerine ücretsiz paylaşılır.

12. Sonuç

7.403 Türkçe sitede laboratuvar ile saha Core Web Vitals ölçümlerinin uyumu, yükleme metriklerinde zayıftan yok denecek düzeye kadar gerilemekte; yalnız düzen kararlılığında orta düzeye çıkmaktadır. Laboratuvar yükleme skorları gerçek-kullanıcı deneyiminin güvenilir bir vekili değildir ve sistematik olarak kötümserdir. Uygulayıcılara önerimiz: kararı sahaya, teşhisi laboratuvara bırakmak. Sonraki çalışmalar farklı throttling yapılandırmalarının ve cihaz sınıflarının uyuma etkisini sınamayı hedefler.

Beyanlar

VERİ ERIŞİLEBİLİRLİĞİ

Anonimleştirilmiş, eşleştirilmiş laboratuvar-saha veri seti ve yeniden-üretilebilir analiz betiği akademik amaçla talep üzerine ücretsiz paylaşılır: akademi@1st.com.tr.

FİNANSMAN

Bu araştırma herhangi bir dış fon almamıştır; 1st.com.tr bünyesinde yürütülmüştür.

ÇIKAR ÇATIŞMASI

Laboratuvar verisi yazarların geliştirdiği 1st.com.tr motoruyla toplanmıştır (şeffaflıkla beyan edilir); saha verisi bağımsız bir kaynaktan (Google CrUX) alınmıştır. Başka çıkar çatışması yoktur.

Kaynaklar

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet, 327(8476), 307-310. [https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8)

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37-46. <https://doi.org/10.1177/001316446002000104>

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159-174. <https://doi.org/10.2307/2529310>

Google. Chrome UX Report (CrUX): origin-level field data. <https://developer.chrome.com/docs/crux>

Google. Lighthouse: automated tool for improving web page quality. <https://developer.chrome.com/docs/lighthouse/overview>

web.dev. Core Web Vitals. <https://web.dev/articles/vitals>

NASIL ATIF YAPILIR

1st.com.tr Araştırma Birimi (2026). Laboratuvar mı, Gerçek Dünya mı? Türkçe Web'de Lighthouse ve CrUX Ölçümlerinin Uyumuna Üzerine Bir Yöntem-Karşılaştırma Çalışması. 1st.com.tr Akademi: Açık Veri Çalışma Kâğıtları, TR-WEB-2026-04. <https://doi.org/10.5281/zenodo.20732713>