

EVALUASI DAN PERBAIKAN DESAIN ANTARMUKA PADA APLIKASI ANDROID MENGGUNAKAN METODE USER EXPERIENCE QUESTIONNAIRE, USABILITY TESTING, DAN HUMAN CENTERED DESIGN (STUDI KASUS: GUNUNG HARTA TRANSPORT SOLUTIONS)

Fito Briantama Setyanto¹, Eriq Muhammad Adams Jonemaro², Mahardeka Tri Ananta³

^{1,2,3}Universitas Brawijaya, Malang

Email: ¹fitobriantama@student.ub.ac.id, ²eriq.adams@ub.ac.id, ³deka@ub.ac.id

*Penulis Korespondensi

(Naskah masuk: 21 Januari 2024, diterima untuk diterbitkan: 10 November 2025)

Abstrak

Moda transportasi bus merupakan moda yang marak digunakan oleh masyarakat karena biayanya yang relatif lebih murah daripada moda transportasi lain. Gunung Harta Transport Solutions (GHTS) merupakan sebuah perusahaan bus yang berlokasi Kota Malang. Evaluasi dilaksanakan pada aplikasi GHTS milik perusahaan ini karena ditemukan desain yang tidak sesuai dengan 10 prinsip *heuristic*. Evaluasi dilakukan menggunakan metode usability testing, SUS, dan UEQ yang dilakukan di website Maze dan pembagian kuisioner. *Usability testing* dilakukan untuk menghitung 3 aspek, yaitu *effectiveness* menggunakan *completion rate*, *efficiency* menggunakan *time based efficiency*, dan *satisfaction* menggunakan SUS. Sedangkan pada UEQ terdapat 6 skala yang dinilai yaitu *attractiveness*, *perspicuity*, *efficiency*, *dependability*, *stimulation*, dan *novelty*. Pada hasil evaluasi desain awal didapatkan nilai pada aspek *effectiveness* dengan angka 99,2%, aspek *efficiency* dengan angka 0,064 *goals/sec*, dan aspek *satisfaction* mendapat rata-rata 39,5 dengan kelas F atau "Poor". Sedangkan pada aspek UEQ didapatkan benchmark dengan predikat "Bad" di keenam skala yang dinilai. Desain perbaikan dirancang guna meningkatkan hasil evaluasi desain awal dan didapatkan peningkatan di segala aspek yang dinilai. Aspek *effectiveness* mendapatkan angka 100%, aspek *efficiency* mendapat angka 0,109 *goals/sec*, dan aspek *satisfaction* mendapat rata-rata 87,625 dengan kelas "B" atau "Excellent". Ditambah pada aspek UEQ terjadi peningkatan yang sangat signifikan yang dimana mendapatkan benchmark dari rata-ratanya adalah "Excellent" di keenam aspek.

Kata kunci: GHTS, usability testing, user experience questionnaire, system usability scale, transportasi, evaluasi, human centered design

EVALUATION AND IMPROVEMENT OF USER INTERFACE DESIGN IN ANDROID APPLICATION USING USER EXPERIENCE QUESTIONNAIRE, USABILITY TESTING, AND HUMAN CENTERED DESIGN (Case Study: Gunung Harta Transport Solutions)

Abstract

Bus transportation is a widely used mode of travel due to its relatively lower cost compared to other modes of transportation. Gunung Harta Transport Solutions (GHTS) is a bus company located in the city of Malang. An evaluation was conducted on the GHTS application due to identified design discrepancies with the 10 heuristic principles. The evaluation utilized usability testing, SUS, and UEQ methods on the Maze website, along with the distribution of questionnaires. Usability testing aimed to calculate three aspects: effectiveness using completion rate, efficiency using time-based efficiency, and satisfaction using SUS. Meanwhile, in the UEQ there are 6 scales that are assessed, namely attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. In the initial design evaluation, effectiveness scored 99.2%, efficiency scored 0.064 goals/second, and satisfaction averaged 39.5 with a grade of F or "Poor." Meanwhile, the UEQ aspect received a benchmark rating of "Bad" across all six scales evaluated. Improvement designs were implemented to enhance the initial design evaluation results, resulting in improvements in all evaluated aspects. Effectiveness scored 100%, efficiency scored 0.109 goals/second, and satisfaction averaged 87.625 with a grade of B or "Excellent." Additionally, there was a significant increase in UEQ aspects, achieving an "Excellent" benchmark across all six scales.

Keywords: GHTS, usability testing, user experience questionnaire, system usability scale, transportation, evaluation, human centered design

1. PENDAHULUAN

Dalam era modern ini, transportasi memiliki peran penting dalam mendukung berbagai aktivitas sehari-hari masyarakat. Transportasi menjadi sarana untuk memindahkan orang dan barang dari satu tempat ke tempat lain (Nova, Widi 2019). Pentingnya transportasi semakin terbantu dengan perkembangan teknologi yang pesat. Hal ini tidak hanya berlaku untuk sektor transportasi, tetapi juga memengaruhi berbagai industri lainnya. Perkembangan teknologi yang terus-menerus mampu mengubah dunia secara menyeluruh dan membentuk cara hidup masyarakat. Oleh karena itu, inovasi terus diperlukan untuk memudahkan kehidupan (Simarmata, 2021).

Salah satu pemanfaatan teknologi yang sangat berperan dalam bidang transportasi adalah pemesanan tiket. Kini, pembelian tiket transportasi seperti pesawat, bus, atau ojek motor dan mobil dapat diakses melalui berbagai aplikasi online. Kemudahan ini memungkinkan pengguna untuk melakukan pemesanan tiket kapan pun dan di mana pun, meningkatkan aksesibilitas dan kenyamanan.

Penelitian ini memfokuskan pada transportasi bus, khususnya aplikasi pemesanan tiket bus dari perusahaan Gunung Harta Transport Solutions. Evaluasi mandiri awal mengungkap beberapa masalah, seperti ketidakhadiran tombol "show password" pada halaman login, warna yang kurang tepat pada pesan kesalahan, dan ketidaktersediaan opsi ubah password tanpa *logout*. Perbaikan diperlukan untuk meningkatkan *user experience* dalam penggunaan aplikasi tersebut.

Permasalahan yang ditemukan, seperti ketidakjelasan tombol "show password" dan kesalahan warna pada feedback login, tidak sesuai dengan prinsip *heuristic* Jakob Nielsen. Terkait informasi yang kurang memadai dan kesalahan warna, tidak memenuhi prinsip "Visibility of system status" dan "Consistency and standards". Selain itu, ketidakefisienan dalam mengganti password yang tidak terintegrasi dalam pengaturan "edit profile" melanggar prinsip "Flexibility and efficiency of use". Perbaikan diperlukan untuk memastikan kesesuaian dengan prinsip-prinsip *heuristic*.

Aplikasi yang baik perlu terus dievaluasi bahkan setelah penelitian selesai, mengingat pentingnya kesesuaian dengan prinsip-prinsip *heuristic* (Prasetyaningtias dkk, 2018). Evaluasi yang berkelanjutan diperlukan untuk mengakomodasi kenyamanan pengguna atau User Experience (UX). UX mencakup persepsi dan respons pengguna setelah menggunakan suatu produk, sistem, atau layanan (ISO 9241-210, 2010), dan perlu dipertimbangkan sejak awal pengembangan produk (Schrepp, 2019).

Metode evaluasi UX yang digunakan dalam penelitian ini melibatkan User Experience Questionnaire (UEQ) dan Usability Testing dengan aspek yang digunakan adalah *effectiveness*, *efficiency*, dan *satisfaction*. Aspek *effectiveness*

menggunakan perhitungan *time completion rate*, aspek *efficiency* menggunakan perhitungan *time based efficiency*, sedangkan aspek *satisfaction* menggunakan kuisioner SUS sebagai sarana pengujiannya. UEQ memberikan data kuantitatif mengenai pengalaman pengguna secara keseluruhan, sementara Usability Testing memberikan data kualitatif mengenai kegunaan dan masalah konkret yang dialami pengguna selama melaksanakan pengujian. Evaluasi terhadap desain awal mengidentifikasi beberapa masalah, seperti ketidakjelasan tombol "show password" dan kesalahan warna pada feedback login, yang tidak sesuai dengan prinsip-prinsip *heuristic*.

Setelah evaluasi desain awal, peneliti akan memberikan rekomendasi desain perbaikan menggunakan pendekatan Human Centered Design (HCD). HCD merupakan metode pengembangan sistem interaktif yang mempertimbangkan kebutuhan pengguna, pengetahuan kegunaan, serta faktor manusia dan ergonomi (ISO 9241-210, 2019). Desain perbaikan akan diuji melalui Usability Testing dan dievaluasi menggunakan UEQ.

Diharapkan, perbaikan desain antarmuka dengan pendekatan HCD dapat meningkatkan pengalaman pengguna dalam menggunakan aplikasi Gunung Harta, menjadikan aplikasi tersebut lebih efektif, efisien, dan memuaskan.

2. METODE PENELITIAN

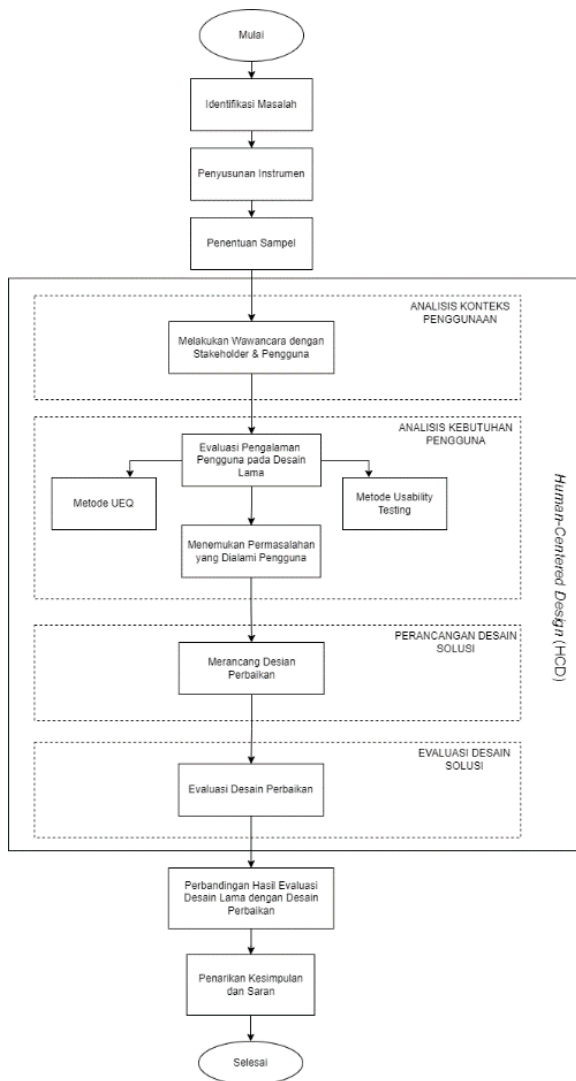
Pada metodologi akan membahas tentang tahapan dari penelitian ini. Tahapan pada penelitian ini bisa dilihat pada gambar 1.

Identifikasi masalah pada penelitian ini dimulai dengan meninjau ulasan di Google Play Store dan melakukan riset mendalam terhadap aplikasi Gunung Harta (GHTS). Tujuannya adalah mengidentifikasi permasalahan yang kemudian disesuaikan dengan 10 prinsip *Heuristic* oleh Jakob Nielsen (1994).

Tahap penyusunan instrumen dilakukan untuk mendapatkan data yang sesuai dengan tahap pengumpulan data yang dilakukan. Untuk menentukan responden yang benar pada penelitian ini, peneliti mengadakan pre-interview kepada calon responden. Setelah itu peneliti juga merancang Google Form yang berisi kuisioner UEQ dan SUS untuk bisa diberikan ke pengguna sesudah melakukan evaluasi.

Tahap penentuan responden ini akan melibatkan 20 responden dari masyarakat umum dengan rentang umur 18 tahun sampai 40 tahun untuk metode User Experience Questionnaire (UEQ) dan penghitungan kuantitatifnya sesuai pedoman Jakob Nielsen (2006), termasuk pengguna yang baru pertama kali menggunakan aplikasi GHTS. Selain itu, ada 5 *representative user* yang akan mengikuti metode usability testing dengan wawancara moderated sesuai pedoman Norman Nielsen (2000). Syarat responden termasuk memiliki kemampuan mengoperasikan

smartphone Android dan pernah memesan tiket GHTS melalui aplikasi.



Gambar 1. Tahapan Penelitian

Tahap selanjutnya adalah melakukan wawancara kepada *stakeholder* dan *representative user* guna mendapatkan kriteria responden yang cocok dan pengumpulan data. Wawancara kepada *stakeholder* dilakukan dengan mendatangi langsung kantor GHTS dan melakukan wawancara dengan karyawan yang ada disana.

Pelaksanaan tahap awal dilakukan untuk bisa mengetahui tingkat *usability* dari desain lama dan Evaluasi kuantitatif menggunakan metode Jakob Nielsen (2006) dengan fokus pada aspek *effectiveness*, *efficiency*, dan *satisfaction*. Evaluasi dilakukan mencakup ke keseluruhan desain yang ada pada GHTS. Penilaian aspek *effectiveness* diambil dari metrik *completion rate*, aspek *efficiency* diambil dari *time based efficiency*, dan aspek *satisfaction* diambil dari pengisian kuisioner SUS. Sedangkan pengambilan data UEQ diambil dari kuisioner

Google Form yang berisikan 26 pertanyaan yang diambil dari UEQ Handbook. Persamaan *completion rate* dan *time based* bisa dilihat pada persamaan berikut.

$$Completion Rate = \frac{Task\ berhasil}{jumlah\ task} \times 100\% \quad (1)$$

$$Time\ based\ efficiency = \frac{\sum_{j=1}^R \sum_{i=1}^N \frac{n_{ij}}{t_{ij}}}{NR} \quad (2)$$

Pelaksanaan evaluasi akan dilakukan di platform Maze dengan metode *moderated* melalui Google Meet. *Task scenario* yang dirancang didapat dari kegiatan pengguna yang bisa dilakukan di aplikasi GHTS. *Task scenario* yang dirancang adalah sebagai berikut.

Tabel 2.1. Task Scenario

Kode	Task Scenario
T1	Melakukan pendaftaran akun
T2	Melakukan login
T3	Memesan tiket bis dari Malang ke Bandung
T4	Mengubah nomor handphone pada akun yang sudah ada
T5	Mengubah password yang sudah ada
T6	Melihat detail cara pembayaran tiket yang sudah dipesan
T7	Mendownload e-ticket yang sudah dipesan

Pada evaluasi tahap awal peneliti akan meminta pendapat mengenai permasalahan yang dialami oleh responden selama mengikuti pengujian. Pengambilan ini akan diambil dari sesi wawancara setelah menjalankan pengujian. Selain itu akan dilakukan analisis sekaligus penghitungan data mengenai 4 penilaian yaitu aspek *effectiveness*, *efficiency*, *satisfaction*, dan aspek UEQ. Setelah melakukan analisis terhadap 4 penilaian tersebut, akan dilakukan list permasalahan yang dialami oleh responden selama pengujian berlangsung. Nantinya identifikasi permasalahan ini akan dijadikan acuan dalam pembuatan desain perbaikan.

Setelah melakukan evaluasi terhadap desain lama, peneliti akan merancang desain perbaikan berdasarkan identifikasi permasalahan desain sebelumnya yang akan dijadikan acuan dalam membuat rancangan desain. Pembuatan desain perbaikan dimulai dari perancangan *sitemaps* yang baru sampai ke *hi-fi prototype* yang akan dibuat menggunakan Figma.

Tahap selanjutnya adalah melakukan evaluasi terhadap desain perbaikan, sama seperti yang dilakukan terhadap desain lama di awal. Tahap ini dilakukan untuk mendapatkan tingkat *usability* dan meminta *feedback* atas desain perbaikan yang sudah dibuat dengan memastikan permasalahan yang dialami responden di evaluasi awal masih dialami atau tidak. Evaluasi dilaksanakan secara *moderated* kepada 20 responden yang sama menggunakan Google Meet dengan mengerjakan *task scenario* yang sama dan pengisian kuisioner.

Selanjutnya adalah tahap membandingkan hasil evaluasi desain awal dengan evaluasi desain perbaikan. Hal ini dilakukan supaya bisa menjawab rumusan masalah dan bisa melihat apakah terdapat perbedaan yang signifikan terhadap rancangan desain yang sudah dibuat. Penghitungan hasil dilakukan menggunakan bantuan *tool* Microsoft Excel. Dilakukan juga uji statistik *t-test* untuk melihat apakah terdapat perbedaan yang signifikan terhadap hasil evaluasi desain awal dengan desain perbaikan. Uji statistik *t-test* dilakukan dengan penghitungan *confidence interval* seperti gambar berikut.

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

Gambar 2.2 Persamaan Confidence Interval

Untuk menghitung *confidence interval* diperlukan nilai rata-rata (\bar{x}) dari seluruh penjumlahan, kemudian nilai *z-score* yaitu 1,9600, nilai standar deviasi, dan akar dari jumlah responden (n).

Tahap terakhir yaitu pengambilan kesimpulan dan pemberian saran yang dilakukan setelah semua langkah penelitian dilakukan.

3. HASIL DAN PEMBAHASAN

Hasil dan pembahasan disini membahas tentang hasil permasalahan yang dialami responden serta perbandingan desain awal dengan desain perbaikan. Selanjutnya akan ada hasil dari evaluasi desain awal dengan desain perbaikan yang nantinya akan dibandingkan

3.1. Permasalahan Responden

Permasalahan responden didapatkan dari wawancara yang akan dirincikan ke beberapa *list*. Permasalahan mencakup kepada keseluruhan desain awal. Rinciannya adalah sebagai berikut.

Tabel 1. Kendala Responden Pada Desain Awal

Kode Kebutuhan	Kendala
GHTSK-1	Halaman setelah splash screen tidak diperlukan karena menambah langkah untuk pengguna login ke aplikasi
GHTSK-2	Tidak hadirnya button "Show Password" pada pengisian password
GHTSK-3	Tidak adanya <i>feedback</i> yang baik saat password dan konfirmasi password sudah sesuai pada halaman daftar
GHTSK-4	Pendaftaran yang masih menggunakan email, disarankan untuk menambah opsi no.telpon
GHTSK-5	<i>Navigation bar</i> lebih baik diletakkan di halaman utama langsung (di bawah layar)
GHTSK-6	Bahasa yang masih bercampur antara bahasa Indonesia dengan bahasa Inggris
GHTSK-7	Warna <i>feedback</i> yang berwarna hijau saat pengguna gagal login
GHTSK-8	Tampilan desain dengan penerapan warna yang tidak konsisten dan terkesan kaku.
GHTSK-9	Pemesanan tiket bisa langsung diakses di halaman utama saja

Kode Kebutuhan	Kendala
GHTSK-10	Peletakkan Ubah password lebih baik ditambahkan di dalam pengaturan <i>user account</i>
GHTSK-11	Button kotak yang berada di overlay untuk pemilihan kursi yang akan dibayar dirasa tidak perlu untuk ditekan
GHTSK-12	Informasi range harga kursi disediakan saat memilih kursi
GHTSK-13	Saat pemilihan kursi, lebih baik disediakan langsung pemilihan hari, sehingga tidak perlu kembali mencari di halaman "Cari Rute"
GHTSK-14	Penambahan opsi pembayaran e-wallet pada pembayaran tiket bus
GHTSK-15	Penelitian rute pada tiket lebih baik dipersingkat dengan tulisan "tampilkan" supaya tampilan tidak bersempitan

3.2. Hasil Evaluasi Aspek *Effectiveness* Menggunakan *Completion Rate* pada Desain Awal

Hasil pada evaluasi ini digolongkan menjadi dua, yaitu berhasil (1) dan gagal (0) sesuai pedoman Jeff Sauro (2011). Pengguna tidak diperbolehkan untuk bertanya, akan tetapi diberikan opsi untuk menyerah.

Tabel 2. Hasil Keberhasilan Task Desain Lama

	T1	T2	T3	T4	T5	T6	T7
R1	S	S	S	S	S	S	S
R2	S	S	S	S	S	S	S
R3	S	S	S	S	S	S	S
R4	S	S	S	S	S	S	S
R5	S	S	S	S	G	S	S
R6	S	S	S	S	S	S	S
R7	S	S	S	S	S	S	S
R8	S	S	S	S	S	S	S
R9	S	S	S	S	S	S	S
R10	S	S	S	S	S	S	S
R11	S	S	S	S	S	S	S
R12	S	S	S	S	S	S	S
R13	S	S	S	S	S	S	S
R14	S	S	S	S	S	S	S
R15	S	S	S	S	S	S	S
R16	S	S	S	S	S	S	S
R17	S	S	S	S	S	S	S
R18	S	S	S	S	S	S	S
R19	S	S	S	S	S	S	S
R20	S	S	S	S	S	S	S

Kemudian dilakukan menggunakan persamaan (2) seperti berikut.

$$Completion Rate = \frac{139}{140} \times 100\%$$

$$Completion Rate = 99,2\%$$

Berdasarkan hasil perhitungan yang didapat pada evaluasi desain lama, keseluruhan jumlah *task* yang diberikan hanya satu responden yang gagal pada *task scenario* mengganti password. Hasil tersebut dapat disimpulkan bahwa responden cukup baik untuk mengoperasikan aplikasi GHTS yang sudah ada.

3.3. Hasil Evaluasi Aspek *Efficiency* Menggunakan *Time Based Efficiency* pada Desain Lama

Aspek *efficiency* dihitung menggunakan metrik *time based efficiency* yang dimana datanya diambil dari hasil pengerjaan Maze. Data dicatat dengan satuan detik. Hasilnya dapat dilihat sebagai berikut.

Tabel 3. Durasi Pengerjaan Task Desain Awal

	T1	T2	T3	T4	T5	T6	T7
R1	41,6	3,3	90,8	9,5	25,0	18,3	14,3
R2	23,9	6,4	75,9	8,8	11,7	15,0	4,9
R3	26,5	3,3	63,1	12,8	18,2	5,5	5,7
R4	16,4	5,2	55,5	9,5	12,7	8,0	4,8
R5	82,6	5,5	163,9	52,9	22,9	29,0	12,2
R6	18,2	2,5	46,8	8,3	10,5	13,3	3,2
R7	28,9	9,4	132,6	29,7	16,5	13,2	7,1
R8	24,0	4,8	82,3	18,4	15,0	10,9	5,7
R9	11,0	3,8	58,3	5,3	10,2	10,4	3,0
R10	22,1	6,1	62,9	9,7	10,6	7,0	5,5
R11	21,6	2,8	77,2	11,4	7,9	14,5	3,8
R12	12,2	2,9	62,3	7,5	12,3	14,1	3,5
R13	22,4	9,1	79,7	9,6	16,9	7,7	4,8
R14	26,5	5,8	74,0	11,8	13,5	6,0	4,6
R15	20,8	4,9	70,7	11,3	13,6	7,4	6,2
R16	19,6	4,5	69,5	9,8	11,7	7,7	4,7
R17	31,2	7,7	89,7	13,6	15,6	20,2	8,9
R18	20,9	3,1	75,6	10,2	22,4	5,3	4,5
R19	36,2	4,1	85,7	119,3	19,2	10,9	14,8
R20	27,0	2,4	50,0	7,0	10,6	5,6	4,6

Hasil yang didapat akan dihitung menggunakan rumus (2). Hasilnya adalah sebagai berikut.

$$Time\ based\ efficiency = \frac{(1,031 + 3,990 + 0,245 + 1,596 + 0,412 + 0,781 + 0,931)}{7 \times 20}$$

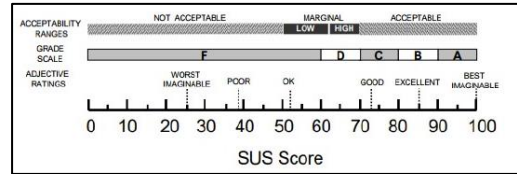
$$Time\ based\ efficiency = \frac{8,989}{140}$$

$$Time\ based\ efficiency = 0,064\ goals/sec$$

Setelah melewati proses perhitungan, didapatkan hasil 0,064 *goals/sec* yang dimana pengguna dapat menyelesaikan setiap tugas dalam 15,625 detik ($1 \div 0,064\ goals/sec$).

3.4. Hasil Evaluasi Aspek *Satisfaction* pada Desain Lama

Pengumpulan data pada aspek *satisfaction* diambil menggunakan kuisioner SUS dan dihitung berdasarkan pedoman dari John Brooke (1986). Data dihitung menggunakan *tool* Excel. Hasil akhir penghitungan SUS akan didapatkan rata-rata dari keseluruhan responden. Hasil akhirnya mendapatkan rata-rata 39,5 yang dimana dapat dikelompokkan ke kelas "F" atau "Poor". Nilai "F" tersebut menandakan bahwa pengguna mengalami permasalahan usability yang serius dan harus segera dibenahi sistemnya guna bisa meningkatkan kualitas usability-nya.



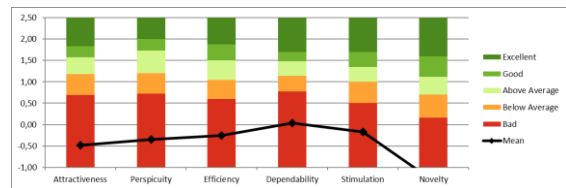
Gambar 3.1. Kelompok Nilai SUS (Bangor, Kortum, dan Miller, 2009)

3.5. Hasil Evaluasi UEQ Pada Desain Lama

Pengumpulan data UEQ diambil dari pengisian kuisioner dan dihitung menggunakan *UEQ Tool Analysis* yang berupa Excel. Hasilnya dapat dilihat pada tabel dan gambar di bawah.

Tabel 4. Rata-rata dan Performa UEQ Desain Lama

Skala	Rata-rata	Perbandingan ke performa
Attractiveness	-0,48	Bad
Perspicuity	-0,35	Bad
Efficiency	-0,25	Bad
Dependability	0,04	Bad
Stimulation	-0,18	Bad
Novelty	-1,31	Bad



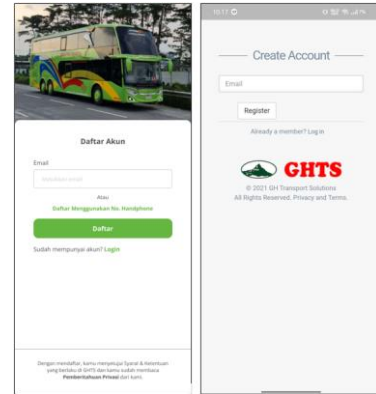
Gambar 2. Benchmark UEQ Desain Lama

Dari hasil yang didapat desain lama mendapatkan hasil yang cukup buruk di keenam skala yang dinilai dengan berada pada level negatif kecuali pada skala *dependability*, dan apabila ingin dikomparasi rata-ratanya dengan *benchmark* sesuai *UEQ tool analysis*, keenam skala tersebut bernilai "Bad". Untuk skala "Attractiveness" atau ketertarikan disini mendapatkan nilai -0,48. Nilai ini dapat diartikan bahwa desain antarmuka yang diterapkan pada desain lama tidak cukup menarik untuk dilihat oleh pengguna. Kemudian pada skala "Perspicuity" atau kejelasan mendapatkan nilai -0,35. Nilai ini juga dapat diartikan bahwa pengguna cukup kesulitan dan sulit dalam beradaptasi dalam menggunakan sistemnya. Selanjutnya skala "Efficiency" atau efisiensi mendapatkan nilai -0,25. Nilai tersebut juga dapat diartikan bahwa pengguna tidak cukup puas dengan keefisienan dari penggunaan sistem aplikasinya. Selanjutnya skala "Dependability" atau ketepatan yang mendapatkan nilai 0,04 yang dimana skala ini bernilai paling bagus diantara semua skala lainnya. Nilai ini dapat diartikan pengguna masih kurang bisa menebak langkah yang disediakan oleh sistem dan kurang bisa mengendalikan interaksi yang ada pada sistem. Sekala selanjutnya adalah "Stimulation" atau stimulasi dengan nilai 0,18. Nilai ini dapat diartikan bahwa desain lama yang diterapkan oleh GHTS beserta usability kurang bisa memotivasi pengguna

untuk memakai produk ini kembali. Skala terakhir adalah “Novelty” atau kebaruan dengan nilai -1,31. Nilai ini merupakan nilai terendah dari semua skala yang dinilai, menandakan bahwa desain yang diterapkan pada aplikasi GHTS sudah ketinggalan zaman dan kurang bisa menyesuaikan dengan desain antarmuka aplikasi yang sudah ada di era sekarang.

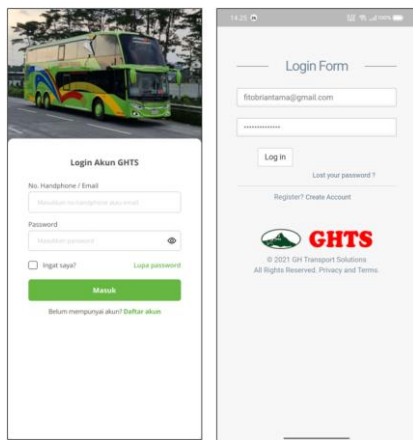
3.6. Perbandingan Desain Perbaikan Dengan Desain Awal

Pada subbab ini akan dijelaskan perbedaan dari desain perbaikan (DP) dengan desain lama (DL) guna menjawab 15 permasalahan responden yang sudah dituliskan.



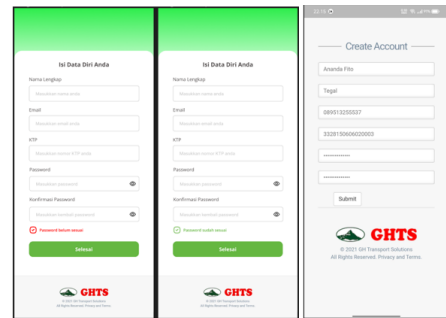
Gambar 3.5. Halaman Daftar Pada DP (a) dan DL(b)

Pada desain perbaikan ditambahkan opsi daftar akun menggunakan untuk pengguna. Kemudian juga tampilan halaman dan tombol diperbaiki.



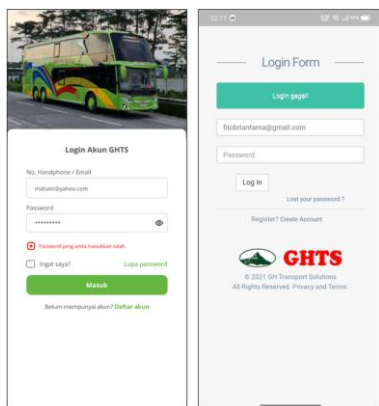
Gambar 3. Halaman Login pada DP (a) dan DL (b)

Terdapat perbedaan tampilan pada DP, yaitu tampilan desain, perubahan warna serta penebalan pada tulisan “lupa password” dan “daftar akun”.



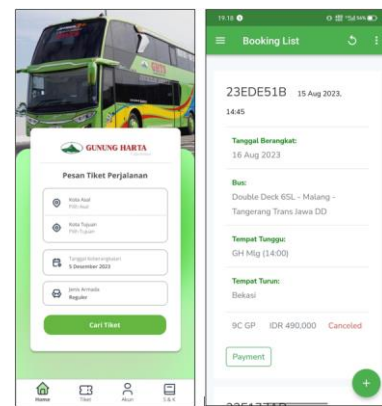
Gambar 6. Feedback Konfirmasi Password Pada DP (a) dan DL (b)

Pada desain perbaikan peneliti menambahkan *feedback* pada pengisian “konfirmasi password” supaya pengguna bisa lebih yakin terhadap password yang diketik.



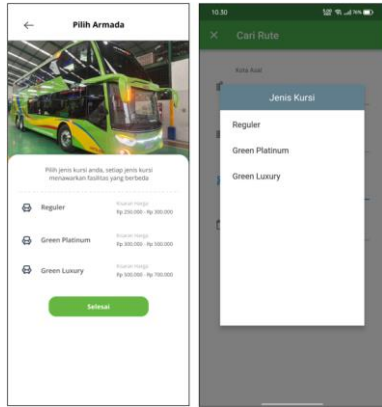
Gambar 4. Feedback Gagal Login Pada DP (a) dan DL (b)

Pada DL, *feedback* yang diterima pengguna masih berwarna hijau, sedangkan pada DP sudah berwarna merah supaya menambah *awareness*. Terdapat juga tombol “show password” pada pengisian password.



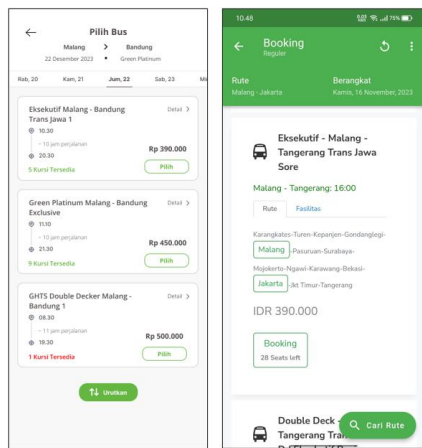
Gambar 7. Halaman Utama Pada DP (a) dan DL (b)

Pada halaman utama kini pengguna bisa mengakses pemesanan tiket secara langsung. Selain itu *navigation bar* pada desain perbaikan dipindahkan ke bawah layar agar pengguna tidak perlu menekan tombol burger di kiri atas untuk mengakses *navigation bar*.



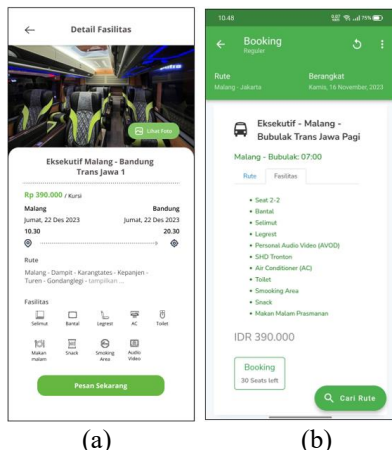
Gambar 8. Pemilihan Armada Pada DP (a) dan DL (b)

Pada DP, pemilihan armada disediakan informasi tentang rentang harga dari masing-masing armada, sehingga pengguna bisa memperkirakan *budget* yang dimilikinya. Kemudian pengguna harus menekan tombol “Selesai” apabila sudah memilih armada yang diinginkan.



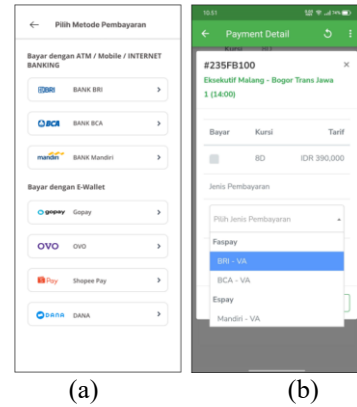
Gambar 9. Daftar Bus Pada DP (a) dan DL (b)

Pada DP, pengguna bisa berganti-ganti tanggal tanpa harus kembali ke halaman sebelumnya dengan menekan tulisan hari di atas. Terdapat juga informasi pemberangkatan dan perkiraan sampai ke kota tujuan.



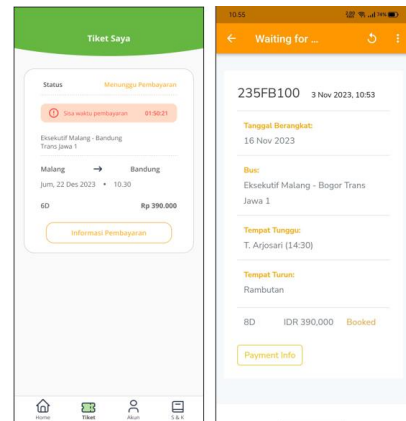
Gambar 10. Detail Fasilitas Pada DP (a) dan DL (b)

Pada DP, pengguna bisa mengakses detail fasilitas apabila menekan kata “detail” pada daftar bus yang tersedia. Tampilan fasilitas pada DP ditambahkan gambar supaya lebih menarik.



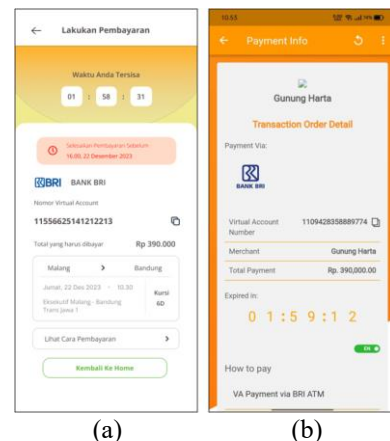
Gambar 11. Metode Pembayaran Pada DP (a) dan DL (b)

Pada DP metode pembayaran ditambahkan dengan jenis *e-wallet* yang ada. Tampilan juga diperbaiki agar lebih nyaman dilihat.



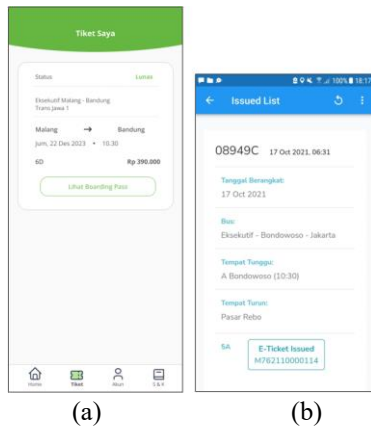
Gambar 3.12. Menunggu Pembayaran Tiket Pada DP (a) dan DL (b)

Pada DP tampilan diperbaiki dan ditambahkan informasi sisa waktu pembayaran dengan warna merah.



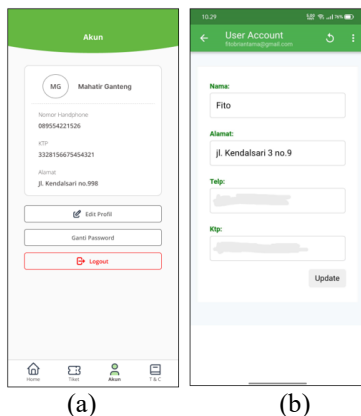
Gambar 13. Informasi Pembayaran Pada DP (a) dan DL (b)

Pada DP tampilan informasi pembayaran diperbaiki. Tombol untuk melihat cara pembayaran juga diperjelas.



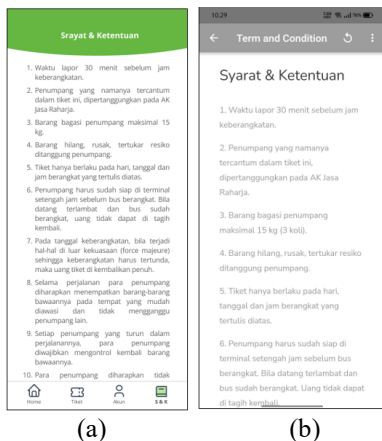
Gambar 14. Tiket Diterbitkan Pada DP (a) dan DL (b)

Pada DP tampilan tiket diterbitkan diubah warnanya menjadi hijau supaya mengikuti warna utama aplikasi. Penggunaan bahasa Indonesia diterapkan juga disini.



Gambar 15. Menu Akun Pada DP (a) dan DL (b)

Pada DP tampilan akun diperbaiki supaya bisa terlihat lebih nyaman. Terdapat juga penambahan fitur “ganti password” supaya pengguna tidak perlu *logout* untuk mengubah password.



Gambar 16. Halaman S&K Pada DP (a) dan DL (b)

Tampilan S&K pada DP hanya terdapat perubahan pada warna navigasi saja mengikuti warna dasar aplikasi. Selebihnya hanya informasi tertulis.

3.7. Perbandingan Hasil Evaluasi Awal dan Evaluasi Desain Perbaikan

Pada evaluasi desain perbaikan dilakukan dengan metode yang sama dengan evaluasi awal. *Task scenario* yang dikerjakan sama, dengan responden yang sama juga.

Pada aspek *effectiveness*, terdapat peningkatan pada desain perbaikan. Pada evaluasi desain awal hasil yang didapat adalah 99,2%, sedangkan hasil yang didapat pada evaluasi desain perbaikan mencapai 100% dan tidak ada yang gagal dalam mengerjakan *task scenario*. Hal ini terjadi dikarenakan yang menjadi permasalahan utama pada desain lama adalah alur dari perubahan password, dan pada desain perbaikan alur dari perubahan password sudah diubah.

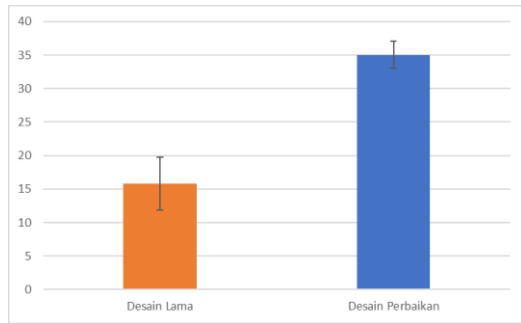
Kemudian pada aspek *efficiency* terdapat peningkatan *time based efficiency* menjadi 0,109 *goals/sec*. Angka tersebut dapat diartikan pengguna mampu menyelesaikan tiap tugas dalam 9,149 detik. Ini merupakan suatu peningkatan yang cukup baik karena pada desain awal didapatkan pengguna mampu menyelesaikan tiap tugas dalam 15,625 detik. Hal ini bisa terjadi karena terdapat beberapa *flow* pada desain perbaikan yang disingkat, seperti pengaksesan halaman tiket dengan segala proses yang dijadikan satu alur, peletakan bar navigasi, perubahan *flow* pada perubahan password pada akun.

Pada aspek *satisfaction* didapatkan penilaian yang sangat baik, yang dimana hasil pada evaluasi desain awal mendapat rata-rata skor SUS 39,5 dengan *grade scale* “F” atau “Poor”, sedangkan pada desain perbaikan mendapatkan rata-rata skor SUS mencapai 87,625 dengan *grade scale* “B” atau bisa dibilang “Excellent”. Dengan rata-rata yang didapatkan, bisa dikatakan bahwa rancangan desain perbaikan yang sekarang bisa lebih diterima oleh pengguna, dan Jeff Sauro (2011) menyebutkan bahwa dengan skor di atas 80,3 pengguna berpotensi untuk merekomendasikan produk kepada koleganya.

Setelah dilakukan penghitungan rata-rata, dilakukan uji statistik t-test untuk melihat perbedaannya. Penghitungan ini dihitung melalui persamaan *confidence interval* seperti gambar 2.2 dan pengaplikasian perhitungannya terdapat pada tabel 3.5.

Desain Lama	Desain Perbaikan
CI	CI
= 15,8	= 35,05
$\pm 1,9600 \frac{8,3766}{\sqrt{20}}$	$\pm 1,9600 \frac{4,3946}{\sqrt{20}}$
CI = 15,8 ± 3,920397	CI = 35,05 ± 2,05677

Nilai *alpha* yang digunakan adalah 95%. Pada desain lama menghasilkan *confidence interval* senilai $15,8 \pm 3,920397$, sedangkan pada desain perbaikan mendapatkan nilai $35,05 \pm 2,056770$. Data tersebut akan disajikan dengan *error bar* seperti gambar berikut.



Gambar 17 Error Bar Pada Usability Score

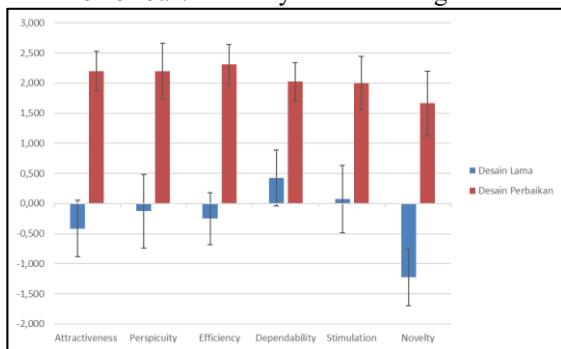
Penghitungan terakhir adalah uji t-test yang didapatkan hasil sebagai berikut.

Tabel 6 Hasil Perhitungan T-test Pada Usability Score

<i>T-test</i>	
<i>Usability Score</i>	0,000000343267431830078

Hasil *t-test* diatas menunjukkan angka yang lebih kecil dari nilai *alpha* yaitu 0,05. Hal ini menunjukkan terdapat peningkatan pada nilai *usability score* antara desain lama dengan desain perbaikan.

Pada penilaian UEQ mendapatkan respon yang sangat positif. Dilakukan uji statistik t-test yang dilakukan pada UEQ *tool analysis* untuk membandingkan hasil evaluasi antara desain lama dengan desain perbaikan yang akan disajikan dalam bentuk *error bar*. Grafiknya adalah sebagai berikut.



Gambar 18 Error Bar Pada Evaluasi UEQ

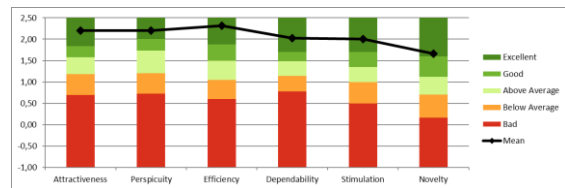
Tabel 7 Selisih Rata-rata UEQ

Skala UEQ	Desain Lama	Desain Perbaikan	Selisih Rata-rata
<i>Attractiveness</i>	-0,483	2,200	2,683
<i>Perspicuity</i>	-0,350	2,200	2,550
<i>Efficiency</i>	-0,250	2,313	2,563
<i>Dependability</i>	0,038	2,025	1,987
<i>Stimulation</i>	-0,175	2,000	2,175
<i>Novelty</i>	-1,313	1,663	2,976

Kemudian dilakukanlah analisis perbandingan dengan melihat selisih rata-rata dari masing-masing skala yang dinilai. Hasilnya seperti tabel 7.

Skala pertama yang dinilai adalah “Attractiveness” atau ketertarikan yang mendapatkan nilai 2,20. Nilai ini dapat diartikan bahwa desain antarmuka yang diterapkan pada desain perbaikan cukup menarik untuk dilihat oleh pengguna. Kemudian pada skala “Perspicuity” atau kejelasan mendapatkan nilai 2,20. Nilai ini juga dapat diartikan bahwa pengguna sudah bisa dengan mudah menyesuaikan dengan lingkungan penggunaan sistemnya, hal ini dikarenakan pada perancangan desain perbaikan, peneliti lebih mengkoordinasi menu dengan tertata, supaya lebih memudahkan penulis. Selanjutnya skala “Efficiency” atau efisiensi mendapatkan nilai 2,31. Nilai tersebut juga dapat diartikan bahwa pengguna cukup puas dengan keefisienan dari penggunaan sistem aplikasinya. Selanjutnya skala “Dependability” atau ketepatan yang mendapatkan nilai 2,03. Nilai ini dapat diartikan bahwa pengguna bisa menebak langkah yang disediakan oleh sistem dan sudah mampu mengendalikan interaksi yang ada pada sistem. Sekala selanjutnya adalah “Stimulation” atau stimulasi dengan nilai 2,00. Nilai ini dapat diartikan bahwa desain perbaikan yang dirancang beserta sudah bisa memotivasi pengguna untuk memakai produk ini kembali. Skala terakhir adalah “Novelty” atau kebaruan dengan nilai 1,66. Nilai ini merupakan nilai terendah dari semua skala yang dinilai, akan tetapi masih mendapatkan predikat *excellent*. Nilai ini menandakan bahwa desain yang diterapkan pada aplikasi GHTS sudah bisa menyesuaikan dengan desain antarmuka aplikasi yang sudah ada di zaman sekarang.

Hasil pada tabel 3.7 merupakan perbedaan dari rata-rata yang didapatkan dari hasil evaluasi desain lama dengan desain perbaikan. Selisih dari keenam aspeknya sangat berjauhan dan selisih yang paling tinggi berada di skala *novelty* (kebaruan). Hal ini terjadi karena peneliti merancang desain perbaikan menyesuaikan dengan desain aplikasi yang sudah ada di zaman sekarang. Keenam skala UEQ pada desain perbaikan mendapatkan predikat “Excellent” apabila disandingkan rata-ratanya dengan benchmark yang ada pada UEQ *handbook*. Berikut hasil dari *benchmark* UEQ pada desain perbaikan.



Gambar 19 Hasil Benchmark UEQ Desain Perbaikan

Setelah melalui fase evaluasi yang dimana dilakukan wawancara kembali mengenai desain perbaikan yang dirancang apakah sudah menjawab permasalahan yang dialami oleh responden pada

desain awal. Didapatkan bahwa semua permasalahan dapat diselesaikan. Terdapat juga masukan dari responden 3 mengenai tulisan “Mendaftar dengan nomor handphone” pada desain perbaikan sebaiknya dibuatkan tombol saja supaya tingkat *awareness* pada pengguna semakin meningkat.

4. KESIMPULAN

Berdasarkan penelitian yang dilakukan didapatkan kesimpulan bahwa desain perbaikan dapat meningkatkan 4 penilaian yang dilakukan, yaitu *effectiveness*, *efficiency*, *satisfaction*, dan Skala UEQ. Temuan 15 permasalahan yang didapat dari hasil wawancara bersama responden juga sudah diatasi pada desain perbaikan menurut pendapat dari responden.

Dimulai dari aspek *effectiveness* mengalami peningkatan dari yang awalnya mendapatkan angka 99,2% pada desain awal, menjadi 100% pada desain perbaikan. Kemudian dari segi *time based efficiency* mengalami peningkatan dari yang awalnya mendapatkan angka 0,064 *goals/sec* (15,625 detik per *task*) menjadi 0,109 *goals/sec* (9,149 detik per *task*). Pada aspek *satisfaction* juga mengalami peningkatan yang sangat tinggi yang dimana pada desain awal mendapatkan rata-rata skor SUS 39,5 dengan *grade scale* “F” atau “Poor”, sedangkan pada desain perbaikan mendapatkan rata-rata skor SUS 87,625 dengan *grade scale* “B” atau “Excellent”. Terakhir pada penilaian UEQ terdapat peningkatan di keseluruhan skala yang dinilai mulai dari *attractiveness*, *perspicuity*, *efficiency*, *dependability*, *stimulation*, dan *novelty*. Keenam aspek tersebut berada pada tingkatan *benchmark* “Excellent”. Peningkatan tertinggi pada skala UEQ ini ada pada skala *novelty* (kebaruan).

Saran penelitian ini diharapkan dapat menjadi sarana pembaruan pada aplikasi GHTS supaya bisa ditingkatkan lagi pada segi desain aplikasinya. Tak lupa juga supaya bisa dilakukan evaluasi secara menerus supaya *user experience* bagi pengguna bisa kian meningkat dan aplikasi GHTS bisa menjadi aplikasi yang lebih baik lagi.

DAFTAR PUSTAKA

U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES., 2014. Usability evaluation basics, 6–7. Retrieved from <https://www.usability.gov/what-and-why/usability-evaluation.html>

- JEFF SAURO, P., 2011. 10 things to know about completion rates. MeasuringU. <https://measuringu.com/completion-rates/>
- BANGOR, A., KORTUM, P. T., & MILLER, J. T., 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3), 114-123.
- LAUGWITZ, B., HELD, T., & SCHREPP, M., 2008. Construction and evaluation of a user experience questionnaire. In *HCI and Usability for Education and Work: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008, Graz, Austria, November 20-21, 2008. Proceedings 4* (pp. 63-76). Springer Berlin Heidelberg.
- NIELSEN NORMAN GROUP., 2012. Usability 101: Introduction to Usability., <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>
- NIELSEN NORMAN GROUP., 2000. World Leaders in Research-Based User Experience. Why You Only Need to Test with 5 Users. Retrieved from <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>
- BROOKE, J., 1996. SUS: A quick and dirty usability scale. 1st ed. s.l.:s.n.
- ISO 9241-11:2018. ISO., 2018. <https://www.iso.org/standard/63500.html>
- RUBIN, J., & CHISNELL, D., 2008. Handbook of usability testing [electronic resource] : How to plan, design, and conduct effective tests (2nd ed.).
- SIMARMATA, J., MANUHUTU, M. A., HERLINAH, H., & SINAMBELA, M., 2021. Pengantar Teknologi Informasi.
- NOVA, D. D. R., & WIDIASTUTI, N., 2019. Pembentukan karakter mandiri anak melalui kegiatan naik transportasi umum. *Comm-Edu (Community Education Journal)*, 2(2), 113-118.
- JEFF SAURO, P. (n.d.), 2011. Measuring usability with the system usability scale (SUS). MeasuringU. <https://measuringu.com/sus/>
- BANGOR, A., KORTUM, P. T., & MILLER, J. T., 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3), 114-123.