

Evaluasi *Usability* Antarmuka Aplikasi Pengerjaan Soal untuk SMA Bina Negara 1 Baleendah Menggunakan *A/B Testing* dan *System Usability Scale*

1st R. Rama Ahmad Waliyya
Sistem Informasi

Telkom University
Bandung, Indonesia

ramawaliyya@student.telkomuniversity
.ac.id

2nd Oktariani Nurul Pratiwi
Sistem Informasi

Telkom University
Bandung, Indonesia

onurulp@telkomuniversity.ac.id

3rd Ilham Perdana
Sistem Informasi

Telkom University
Bandung, Indonesia

ilhamp@telkomuniversity.ac.id

Abstrak — Perkembangan teknologi digital telah mendorong transformasi dalam sistem pendidikan, termasuk dalam proses evaluasi pembelajaran. SMA Bina Negara 1 Baleendah sebelumnya menggunakan Google Form dan Microsoft Excel untuk distribusi soal dan rekap nilai, namun metode tersebut dinilai kurang efisien. Penelitian ini bertujuan mengevaluasi desain antarmuka aplikasi pengerjaan soal berbasis web menggunakan metode *A/B Testing* dan *System Usability Scale* (SUS). *A/B Testing* digunakan untuk mengetahui preferensi visual pengguna terhadap dua versi desain, sedangkan SUS digunakan untuk mengukur *usability* aplikasi berdasarkan persepsi pengguna. Sebanyak 7 guru dan 38 siswa dilibatkan sebagai responden dalam penelitian ini. Hasil *A/B Testing* menunjukkan bahwa guru lebih menyukai desain B pada halaman bank soal, pembuatan kuis, dan pembuatan soal, sedangkan siswa memilih desain B untuk halaman pengerjaan kuis. Rekomendasi perbaikan desain diberikan berdasarkan masukan dari pengguna. Hasil evaluasi SUS menunjukkan skor rata-rata 85 untuk guru dan 88 untuk siswa, yang termasuk kategori “*Best Imaginable*”, menandakan tingkat *usability* yang sangat tinggi. Dengan demikian, sistem yang dikembangkan dinilai telah memenuhi standar kemudahan penggunaan, meskipun masih diperlukan penyempurnaan minor guna meningkatkan pengalaman mengerjakan soal secara daring.

Kata kunci— *A/B Testing*, *Evaluasi UI/UX*, *Pendidikan Digital*, *System Usability Scale*, *Usability*

I. PENDAHULUAN

Di era digital saat ini, transformasi teknologi telah memberikan dampak besar pada dunia pendidikan, termasuk dalam hal sistem evaluasi pembelajaran. Pengalaman pengguna menjadi salah satu aspek penting dalam pengembangan sistem pendidikan digital karena berpengaruh langsung terhadap kenyamanan dan efektivitas pengguna dalam menggunakan aplikasi [1]. SMA Bina Negara 1 Baleendah, sebagai institusi pendidikan menengah, sebelumnya mengandalkan platform seperti Google Form untuk mendistribusikan soal dan Microsoft Excel untuk merekap nilai siswa. Namun, pendekatan tersebut memiliki

banyak keterbatasan, terutama dari sisi efisiensi dan integrasi data [2].

Untuk mengatasi permasalahan tersebut, dikembangkanlah aplikasi berbasis *website* yang memungkinkan guru mendistribusikan soal dan merekap nilai secara otomatis, serta siswa dapat mengerjakan kuis secara daring. Agar desain antarmuka aplikasi dapat memberikan kenyamanan dan kemudahan bagi penggunanya, diperlukan proses evaluasi *usability* yang komprehensif [3]. Evaluasi dilakukan sejak tahap pengembangan awal untuk memastikan antarmuka sistem benar-benar sesuai dengan kebutuhan pengguna, sekaligus menghindari kesalahan desain yang berakibat pada rendahnya adopsi dan kepuasan pengguna. Oleh karena itu, dalam penelitian ini diterapkan metode *A/B Testing* untuk mengetahui preferensi pengguna terhadap tampilan antarmuka [4], serta *System Usability Scale* (SUS) untuk mengukur tingkat *usability* secara kuantitatif dan terstandardisasi [5].

II. KAJIAN TEORI

Kajian teori dalam penelitian ini membahas berbagai konsep yang mendasari proses evaluasi *usability* dan pengalaman pengguna (*User Experience*) pada aplikasi *website* pengerjaan soal di SMA Bina Negara 1 Baleendah. Pembahasan meliputi teori mengenai *usability* dalam konteks desain antarmuka pengguna (UI), pendekatan UX dalam pengembangan sistem berbasis *web* untuk pendidikan, serta metode evaluasi yang digunakan, yaitu *A/B Testing* dan *System Usability Scale* (SUS). Pemahaman terhadap konsep-konsep tersebut menjadi landasan untuk menilai sejauh mana efektivitas desain yang dikembangkan serta mengidentifikasi area yang perlu diperbaiki guna meningkatkan kenyamanan dan kepuasan pengguna dalam menggunakan aplikasi.

A. Usability Testing

Usability Testing merupakan proses evaluasi yang melibatkan pengguna secara langsung untuk berinteraksi dengan produk atau prototipe, dengan tujuan mengamati dan

menemukan kendala dalam penggunaan [6]. Proses ini berperan dalam menemukan kendala terkait kemudahan penggunaan dan memberikan masukan untuk perbaikan desain. Pada akhirnya, hal ini bertujuan untuk meningkatkan efektivitas produk serta kepuasan pengguna. Hasil evaluasi akan menjadi acuan bagi pengembang dalam menyempurnakan produk agar sesuai dengan kebutuhan dan ekspektasi pengguna [7].

Terdapat jenis *usability testing*, yaitu Formatif, dan Summatif [8]. Formatif berfokus pada pilihan desain umum sehingga sering menggunakan prototipe dengan fidelitas rendah. Formatif juga berfokus pada pengumpulan data kualitatif. Sedangkan Summatif mengevaluasi efektivitas dari desain tertentu. Summatif berfokus pada pengumpulan data kuantitatif untuk menilai seberapa efisien dari hasil prototipe.

B. System Usability Scale

System Usability Scale (SUS) adalah alat yang dirancang untuk melakukan evaluasi dari produk, termasuk juga produk digital dan situs *web*. SUS diperkenalkan oleh John Brooke pada tahun 1986. Metode ini sederhana dan dapat diandalkan untuk mengevaluasi kegunaan produk [9].

SUS mencakup 10 pernyataan yang dinilai oleh pengguna dan menghasilkan skor dalam rentang 1 hingga 100, yang digunakan untuk mengevaluasi pengalaman pengguna terhadap sebuah produk [9]. Penerapan dari metode SUS ini sudah sangat luas. Berawal dari sistem kantor elektronik hingga saat ini SUS dapat menilai situs *web* atau produk digital [10]. Adapun daftar pertanyaan SUS yang dapat dilihat pada tabel berikut:

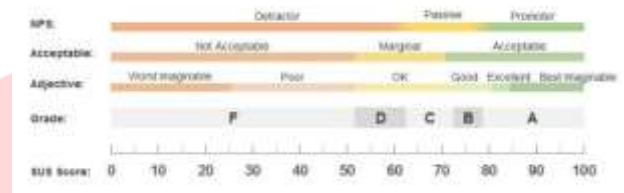
TABEL 1
Daftar Pertanyaan System Usability Scale

No	Pertanyaan
1	Saya merasa akan sering menggunakan sistem ini.
2	Saya merasa sistem ini tidak rumit (mudah digunakan).
3	Saya merasa sistem ini mudah digunakan tanpa bantuan orang lain.
4	Saya merasa berbagai fitur dalam sistem ini saling terintegrasi dengan baik.
5	Saya merasa sistem ini memiliki terlalu banyak inkonsistensi.
6	Saya merasa kebanyakan orang akan bisa cepat belajar menggunakan sistem ini.
7	Saya merasa sistem ini sangat membingungkan untuk digunakan.
8	Saya merasa percaya diri saat menggunakan sistem ini.
9	Saya merasa perlu belajar banyak hal sebelum bisa menggunakan sistem ini.
10	Saya merasa fitur yang tersedia cukup lengkap untuk kebutuhan saya.

Dalam metode ini, untuk pertanyaan yang bersifat positif seperti pada pertanyaan 1, 2, 3, 4, 6, 8, dan 10, skor dihitung dengan mengurangi angka 1 dari jawaban responden (rumus: $n - 1$). Sedangkan untuk pertanyaan yang bersifat negatif seperti pada pertanyaan 5, 7, dan 9, perhitungannya dibalik, yaitu 5 dikurangi jawaban responden (rumus: $5 - n$). Setelah skor dari semua pertanyaan dihitung, nilai-nilai tersebut

dijumlahkan dan kemudian dikalikan dengan angka 2,5, sehingga menghasilkan skor akhir SUS dalam rentang 0 hingga 100 untuk masing-masing responden. Skor individu ini kemudian dijumlahkan dan dirata-rata untuk memperoleh nilai akhir SUS secara keseluruhan dari seluruh responden.

Skor rata-rata tersebut menjadi dasar dalam menilai tingkat kelayakan dan kemudahan penggunaan aplikasi, dan akan dibandingkan dengan skala interpretasi SUS untuk menentukan kategori kualitas antarmuka, apakah masuk kategori *Acceptable*, *Marginal*, atau *Best Imaginable*. Hasil akhir dari pengujian ini dapat dilihat melalui skala interpretasi SUS yang ditampilkan pada Gambar 1 berikut:



GAMBAR 1
Skala Interpretasi System Usability Scale

C. A/B Testing

A/B Testing merupakan metode eksperimen yang digunakan untuk membandingkan dua versi dari suatu elemen seperti tampilan halaman *web*, desain antarmuka, atau fitur aplikasi dengan tujuan mengetahui versi mana yang memberikan performa lebih baik berdasarkan data dan metrik yang telah ditentukan sebelumnya. Metode ini membantu mengidentifikasi elemen yang paling efektif dalam meningkatkan kinerja atau respons pengguna. Versi A biasanya dianggap sebagai kontrol (versi asli), sedangkan versi B adalah variasi yang diuji [11]. Pengujian A/B lebih berfokus pada pengujian algoritma, elemen visual, dan alur kerja karena elemen-elemen ini dapat meningkatkan pengalaman pengguna dan mengoptimalkan kinerja [4].

Pengujian A/B memiliki keuntungan seperti membantu menghilangkan bias subjektif, dapat mengurangi risiko dengan menguji ide sebelum implementasi penuh, serta memiliki iterasi yang cepat untuk menyempurnakan pengalaman pengguna [12]. Namun kekurangan yang dimiliki dari pengujian A/B adalah membutuhkan waktu yang panjang untuk mendapat sampel yang besar, dan sulit untuk mengisolasi pengaruh satu variabel jika terlalu banyak elemen berubah.

III. METODE

Pada penelitian ini dilakukan evaluasi *usability* terhadap prototipe aplikasi *website* pengerjaan soal untuk mengukur sejauh mana kemudahan penggunaan serta kenyamanan pengguna (UX) dalam berinteraksi dengan sistem. Pengujian dilakukan menggunakan metode *A/B Testing* untuk mengetahui preferensi visual pengguna, serta metode *System Usability Scale* (SUS) guna menilai aspek *usability* dan mengidentifikasi elemen desain yang perlu ditingkatkan. Tahap metodologi penelitian ini disajikan dalam gambar 1 berikut:



GAMBAR 2
Metodologi Penelitian

A. Pendahuluan

Bab ini membahas metodologi penelitian yang digunakan dalam proses pengujian dan validasi *Minimum Viable Product* (MVP) dari prototipe aplikasi *website* pengerjaan soal. Penelitian ini menerapkan metode pengujian usability dengan teknik *A/B Testing* dan *System Usability Scale* (SUS) untuk mengevaluasi pengalaman pengguna terhadap tampilan dan kemudahan penggunaan sistem.

B. Studi Literatur

Penelitian ini diawali dengan studi literatur untuk memahami konsep *User Experience* (UX), metode evaluasi *usability*, serta penerapan *A/B Testing* dan *System Usability Scale* (SUS) dalam pengujian desain antarmuka. Studi ini bertujuan untuk menemukan pendekatan yang tepat dalam mengoptimalkan UX pada prototipe aplikasi *website* pengerjaan soal di SMA Bina Negara 1 Baleendah.

C. Metodologi Penelitian

Penelitian ini menggunakan pendekatan *usability testing* untuk mengevaluasi desain antarmuka aplikasi *website* pengerjaan soal. Evaluasi dilakukan dengan dua metode, yaitu *A/B Testing* untuk mengetahui preferensi pengguna terhadap desain visual, dan *System Usability Scale* (SUS) untuk menilai tingkat kemudahan penggunaan sistem.

1. *A/B Testing*, Responden diminta memilih salah satu dari dua desain antarmuka (desain A atau desain B) berdasarkan elemen visual seperti menu kelas, bank soal, halaman kuis, dan rekap nilai, guna mengetahui desain mana yang lebih disukai dan sesuai ekspektasi pengguna.
2. *System Usability Scale*, Setelah implementasi desain pilihan, pengguna diminta menggunakan aplikasi

dan mengisi kuesioner SUS yang terdiri dari 10 pernyataan untuk mengukur usability secara kuantitatif.

D. Menentukan Kebutuhan Penelitian

Sebelum melakukan pengujian, peneliti membutuhkan desain B yang akan diuji dan kuisisioner SUS untuk pengujian *System Usability Scale* (SUS).

1. Desain B

Desain B dikembangkan berdasarkan hipotesis desain untuk meningkatkan efisiensi dan kenyamanan pengguna dalam mengerjakan soal. Berikut adalah perbandingan desain A dan B yang akan digunakan untuk pengujian A/B:

TABEL 2
Desain B Halaman Guru

Fitur	Desain B
Halaman Kelas	
Bank Soal	
Membuat Bank Soal	
Membuat Soal	

Membuat Kuis	
Halaman Nilai	

Halaman Preview Kuis	
Halaman Nilai	

TABEL 3
Desain B Halaman Siswa

Fitur	Desain B
Dashboard Siswa	
Halaman Kelas	
Halaman Pengerjaan Soal	

2. Kuisiener System Usability Scale

Kuesiener System Usability Scale (SUS) dimanfaatkan untuk mengevaluasi tingkat *usability* dari prototipe berdasarkan pengalaman pengguna. Setiap partisipan akan diminta untuk mengisi kuesiener setelah mereka menyelesaikan skenario tugas yang telah ditentukan.

E. Mengumpulkan Data

Pengumpulan data dalam penelitian ini dilakukan melalui pengujian *usability testing* dengan melibatkan responden yang sesuai dengan karakteristik pengguna sistem, yaitu guru dan siswa SMA Bina Negara 1 Baleendah. Tahapan pengumpulan data terdiri dari:

- Penentuan Responden – Pengujian dilakukan dengan melibatkan sekitar 5–10 guru dan 20–40 siswa sebagai partisipan.
- Pelaksanaan Pengujian A/B – Partisipan diminta memilih versi desain antarmuka (A atau B) yang lebih sesuai dengan preferensi mereka berdasarkan visualisasi mockup statis.
- Penyebaran Kuesiener SUS – Setelah versi final dari antarmuka dipilih dan diimplementasikan, responden diminta untuk mengisi kuesiener System Usability Scale (SUS) guna mengevaluasi tingkat kemudahan penggunaan sistem.

F. Analisis Data

Data yang terkumpul dari proses pengujian kemudian dianalisis guna memperoleh pemahaman terkait aspek *usability* aplikasi pengerjaan soal. Analisis dilakukan melalui dua pendekatan sebagai berikut:

- Analisis Preferensi A/B Testing – Data dari pemilihan desain antarmuka versi A dan B dianalisis untuk mengetahui kecenderungan visual pengguna serta elemen desain yang paling disukai oleh guru dan siswa.
- Analisis Kuesiener SUS – Hasil pengisian kuesiener System Usability Scale oleh responden dihitung

untuk memperoleh skor akhir, yang memberikan gambaran kuantitatif mengenai tingkat kemudahan penggunaan aplikasi berdasarkan persepsi pengguna.

G. Kesimpulan

Pada bagian kesimpulan, akan dipaparkan hasil analisis dan rekomendasi berdasarkan sintesis dari metode evaluasi *A/B Testing* dan kuesioner *System Usability Scale* (SUS) yang telah dianalisis sebelumnya. Evaluasi dengan kedua metode ini memungkinkan peneliti untuk mengukur tingkat kepuasan dan persepsi pengguna terhadap aspek usability dari aplikasi *website* pengerjaan soal di SMA Bina Negara 1 Baleendah. Berdasarkan temuan yang diperoleh, beberapa rekomendasi perbaikan disusun guna meningkatkan kualitas tampilan antarmuka serta pengalaman pengguna. Selain itu, hasil dari penelitian ini diharapkan dapat menjadi acuan dalam pengembangan sistem serupa di bidang pendidikan, khususnya dalam penerapan desain UI/UX yang lebih efisien dan responsif terhadap kebutuhan pengguna.

IV. HASIL DAN PEMBAHASAN

A. Pengujian A/B

A/B Testing dilakukan dengan membandingkan dua versi desain antarmuka, yaitu desain A dan desain B, guna mengetahui preferensi pengguna terhadap tampilan visual aplikasi pengerjaan soal. Total responden pada guru adalah 7 responden sedangkan untuk siswa berjumlah 37 responden. Berikut adalah hasil pengujian A/B yang disajikan dalam bentuk tabel:

TABEL 4
Hasil Pengujian A/B Guru

Desain	Desain A	Desain B
Menu Kelas	6	1
Bank soal	1	6
Membuat Bank Soal	1	6
Membuat Soal	1	6
Membuat Kuis	0	7
Melihat Nilai Siswa	5	2
Total	14	28

TABEL 5
Hasil Pengujian A/B Siswa

Desain	Desain A	Desain B
Dashboard Siswa	32	6
Menu Kelas	32	6
Pengerjaan Kuis	0	38
Preview Kuis	29	9
Melihat Nilai Siswa	29	9
Total	122	68

Berdasarkan Tabel 4, hasil pengujian terhadap guru menunjukkan bahwa Desain B lebih dominan dipilih oleh responden dibandingkan Desain A. Dari total 42 pilihan, Desain B memperoleh 28 suara, sedangkan Desain A hanya memperoleh 14 suara. Desain B lebih disukai pada hampir semua bagian, terutama pada halaman “Membuat Kuis” dan “Membuat Soal” yang masing-masing dipilih oleh 7 dan 6 responden. Sementara itu, Desain A hanya lebih unggul dalam bagian “Melihat Nilai Siswa”.

Sedangkan pada pengujian terhadap siswa pada Tabel 5, hasil menunjukkan kecenderungan yang lebih beragam dan kontras. Desain A lebih banyak dipilih untuk halaman

Dashboard Siswa dan Menu Kelas, dengan masing-masing mendapatkan 32 suara dari total 38 responden. Namun, pada bagian Pengerjaan Kuis, Desain B mendominasi penuh dengan 38 suara, tanpa satupun memilih Desain A. Hal serupa juga terjadi pada fitur “Preview Kuis” dan “Melihat Nilai Siswa”, di mana Desain A tetap unggul, namun tidak terlalu jauh dari Desain B. Secara total, Desain A mendapatkan 122 suara, sedangkan Desain B memperoleh 68 suara dari siswa.

Berdasarkan hasil pengujian A/B pada guru dan siswa, hasil pemilihan desain untuk meningkatkan nilai pengalaman pengguna pada aplikasi pengerjaan soal ini disajikan dalam tabel 6 berikut:

TABEL 6
Rekomendasi Perbaikan Desain

No	Peran	Pilihan Desain	Alasan
1	Guru	Desain B Halaman Bank Soal	Karena terdapat fitur tag yang membantu dalam mengelompokkan bank soal
2	Guru	Desain B Membuat Bank Soal	Karena terdapat fitur tag yang membantu dalam mengelompokkan bank soal
3	Guru	Desain B Membuat Kuis	Karena jika guru mempunyai banyak bank soal, layout tile akan mempermudah dalam mencari bank soal
4	Guru	Desain B Membuat Soal	toggle required tidak perlu ditambahkan karena siswa memiliki kewajiban untuk menjawab pertanyaan yang diberikan.
5	Siswa	Desain B Pengerjaan Kuis	Layout nomor yang sedikit akan lebih enak dilihat daripada memaksakan

B. Pengujian System Usability Scale

Pengujian ini melibatkan guru dan siswa dengan 7 responden dari guru dan 38 responden dari siswa. Masing-masing akan diberi label G1-G7 untuk guru dan S1-S38 untuk siswa. Data hasil pengisian kuesioner oleh para responden kemudian direkap dan disajikan dalam bentuk tabel, sebagaimana tercantum pada Tabel 7 berikut ini:

TABEL 7
Hasil Perhitungan System Usability Scale Guru

Responden	HASIL KUESIONER										SUS
	Q 1	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8	Q 9	Q 10	
G1	3	3	2	4	1	3	3	4	3	4	75
G2	4	3	4	4	2	3	3	4	4	4	87.5
G3	4	4	4	4	3	2	3	4	4	4	90
G4	3	3	2	4	4	3	3	4	3	4	82.5

G5	4	2	2	4	4	3	3	4	3	4	82.5
G6	4	3	3	3	3	4	4	4	4	3	87.5
G7	4	3	3	3	4	3	4	4	4	4	90
RATA-RATA											85

TABEL 8
Hasil Perhitungan System Usability Scale Siswa

Respon den	HASIL KUISIONER										SUS
	Q 1	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8	Q 9	Q 10	
S1	4	4	4	5	2	5	2	4	2	5	82.5
S2	5	4	4	5	1	5	1	4	1	5	92.5
S3	5	3	3	5	1	3	1	4	1	5	82.5
S4	5	4	5	4	2	4	2	4	2	5	82.5
S5	4	4	4	4	1	4	1	4	1	5	85
S6	5	4	3	5	1	3	1	5	3	5	82.5
S7	5	4	4	4	2	5	1	5	1	4	87.5
S8	3	4	4	5	1	4	1	4	1	4	82.5
S9	5	5	5	5	1	5	1	5	1	5	100
S10	5	3	3	4	2	3	3	5	3	5	70
S11	5	4	4	5	2	4	1	4	1	5	87.5
S12	4	4	4	5	2	4	1	4	1	4	82.5
S13	5	4	4	5	1	5	2	5	2	5	90
S14	5	5	5	5	2	4	2	5	2	5	90
S15	5	5	5	5	1	5	1	5	1	5	100
S16	4	4	4	4	1	4	1	4	1	4	82.5
S17	5	4	4	4	2	4	2	4	2	5	80
S18	4	5	5	5	2	5	1	5	1	5	95
S19	5	4	1	5	1	4	1	5	1	5	85
S20	5	4	4	5	1	4	1	5	1	5	92.5
S21	5	5	5	5	1	4	1	5	1	5	97.5
S22	5	4	4	5	1	5	1	5	1	5	95
S23	5	5	5	5	1	5	1	5	1	5	100
S24	5	4	4	5	1	5	1	5	1	5	95
S25	5	4	4	5	1	3	1	5	1	5	90
S26	5	5	4	4	1	4	1	5	1	5	92.5
S27	5	5	4	5	1	5	1	5	1	5	97.5
S28	5	4	4	4	1	4	1	4	1	4	85
S29	4	4	4	4	1	4	1	4	1	4	82.5
S30	3	4	4	4	1	4	1	4	1	4	80
S31	5	5	5	4	1	3	1	5	1	4	90
S32	5	5	4	5	1	4	2	4	2	5	87.5

S33	3	4	4	5	2	3	2	5	2	5	77.5
S34	5	4	3	4	1	5	1	4	1	5	87.5
S35	5	5	4	5	1	5	1	4	1	5	95
S36	5	4	4	5	4	4	1	4	3	5	77.5
S37	5	5	4	4	2	5	1	5	2	5	90
S38	4	5	5	5	1	4	2	5	2	5	90
RATA-RATA											88

Pada tahap pelaksanaannya, responden diminta memberikan penilaian terhadap sepuluh pernyataan dalam kuesioner SUS menggunakan skala Likert 1 hingga 5. Berdasarkan hasil pengisian, diperoleh rata-rata skor SUS sebesar 85 dari 7 guru, dan 88 dari 38 siswa. Kedua skor tersebut berada dalam kategori “*Best Imaginable*” menurut interpretasi skala SUS yang berada pada rentang 84,1 hingga 100. Temuan ini menunjukkan bahwa tingkat keberterimaan (acceptability) aplikasi sangat tinggi di mata kedua kelompok pengguna utama.

Tingginya skor yang diperoleh juga mencerminkan bahwa desain antarmuka aplikasi sudah dianggap mudah digunakan, nyaman, dan memenuhi kebutuhan pengguna dalam mengerjakan soal secara mandiri. Selain itu, kecenderungan pengguna terhadap aplikasi ini berada pada tingkat “*Promoter*”, yang berarti mereka berpotensi merekomendasikan aplikasi kepada orang lain karena pengalaman pengguna yang positif.

V. KESIMPULAN

Penelitian ini bertujuan untuk mengevaluasi dan meningkatkan pengalaman pengguna pada aplikasi *website* pengerjaan soal di SMA Bina Negara 1 Baleendah melalui pengujian *A/B Testing* dan *System Usability Scale* (SUS). Hasil *A/B Testing* menunjukkan bahwa sebagian responden lebih menyukai desain antarmuka versi B di beberapa halaman seperti halaman bank soal, membuat bank soal, membuat kuis, dan membuat soal pada peran guru. Selain itu, siswa lebih menyukai desain B pada pengerjaan kuis karena dinilai lebih rapi dan memudahkan navigasi. Oleh karena itu, terdapat rekomendasi perbaikan desain sesuai preferensi guru dan siswa.

Di sisi lain, hasil evaluasi menggunakan SUS menghasilkan skor rata-rata 85 untuk guru dan 88 untuk siswa, yang keduanya termasuk dalam kategori “*Best Imaginable*”. Skor ini menunjukkan tingkat usability yang sangat tinggi serta potensi besar bagi pengguna untuk merekomendasikan aplikasi. Dengan demikian, dapat disimpulkan bahwa desain aplikasi telah memenuhi standar kemudahan penggunaan yang baik, meskipun tetap dibutuhkan penyempurnaan minor untuk mendukung pengalaman mengerjakan soal secara daring yang lebih optimal.

REFERENSI

- [1] Muhammad Iqbal Nugraha, Rudi Kurniawan, and Tati Suprpti, “System Usability Scale in Optimizing the Usability Value of the Tutoring and Private Lessons

- Ayo Pintar Application in Cirebon City,” *j. of artif. intell. and eng. appl.*, vol. 4, no. 2, pp. 1471–1475, Feb. 2025, doi: 10.59934/jaiea.v4i2.930.
- [2] M. L. Jundillah, R. Ramadiani, and A. Azainil, “Measuring the quality of STAR websites using the System Usability Scale (SUS) method,” *TEPIAN*, vol. 6, no. 1, pp. 45–49, Mar. 2025, doi: 10.51967/tepian.v6i1.3201.
- [3] I. M. A. Sudestra, N. W. E. Agustini, I. M. A. O. Gunawan, G. Indrawan, and M. Hakimi, “IMPROVING DIGITAL LEARNING: EVALUATING THE U LEARN LMS WITH THE SYSTEM USABILITY SCALE,” *jipi. jurnal. ilmiah. penelitian. dan. pembelajaran. informatika.*, vol. 9, no. 4, pp. 2325–2332, Nov. 2024, doi: 10.29100/jipi.v9i4.6910.
- [4] F. Quin, D. Weyns, M. Galster, and C. C. Silva, “A/B Testing: A Systematic Literature Review,” 2023, *arXiv*. doi: 10.48550/ARXIV.2308.04929.
- [5] M. Hyzy *et al.*, “System Usability Scale Benchmarking for Digital Health Apps: Meta-analysis,” *JMIR Mhealth Uhealth*, vol. 10, no. 8, p. e37290, Aug. 2022, doi: 10.2196/37290.
- [6] M. Hertzum, *Usability Testing: A Practitioner’s Guide to Evaluating the User Experience*. in Synthesis Lectures on Human-Centered Informatics. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-031-02227-2.
- [7] S. Kravchenko, G. Marchuk, T. Loktikova, Y. Grishkun, and State University «Zhytomyr polytechnic», “USABILITY TESTING METHODS FOR ASSESSING A MOBILE APPLICATION,” *Herald KhmNU.TS*, vol. 317, no. 1, pp. 111–118, Feb. 2023, doi: 10.31891/2307-5732-2023-317-1-111-117.
- [8] J. Lazar, J. H. Feng, and H. Hochheiser, “Usability testing,” in *Research Methods in Human Computer Interaction*, Elsevier, 2017, pp. 263–298. doi: 10.1016/B978-0-12-805390-4.00010-8.
- [9] I. M. Herawati and D. Azahra, “EVALUASI USABILITY WEBSITE JASUDA.NET MENGGUNAKAN SYSTEM USABILITY SCALE (SUS),” *jipi. jurnal. ilmiah. penelitian. dan. pembelajaran. informatika.*, vol. 9, no. 2, pp. 994–1000, Jun. 2024, doi: 10.29100/jipi.v9i2.4328.
- [10] A. Nuswantari, Y.-T. Wu, and H. D. Surjono, “System Usability Scale Measurement on Synhchronous Online Argumentation Learning System,” in *Proceedings of the International Conference on Online and Blended Learning 2019 (ICOBL 2019)*, Yogyakarta, Indonesia: Atlantis Press, 2020. doi: 10.2991/assehr.k.200521.031.
- [11] R. Kohavi and R. Longbotham, “Online Controlled Experiments and A/B Tests,” in *Encyclopedia of Machine Learning and Data Science*, D. Phung, G. I. Webb, and C. Sammut, Eds., New York, NY: Springer US, 2023, pp. 1–13. doi: 10.1007/978-1-4899-7502-7_891-2.
- [12] R. Kohavi, D. Tang, and Y. Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*, 1st ed. Cambridge University Press, 2020. doi: 10.1017/9781108653985.