Evaluasi Usabilitas Mobile Learning Matematika SMA Menggunakan SUS dan Evaluasi Heuristik

Daniel Azarya Tafuama[1], Jeannette Kesia Naomi Liow[2], Keefa Noah Lasut[3], Ade Yusupa[4], Henry Valentino Florensius Kainde[5], Sary Paturusi[6]

[1,2,3,4,5] Program Studi Informatika, Universitas Sam Ratulangi, Manado

Email korespondensi: danielazarya026@student.unsrat.ac.id, jeannetteliow026@student.unsrat.ac.id, keefalasut026@student.unsrat.ac.id, ade@unsrat.ac.id, valentkainde@unsrat.ac.id, sarypaturusi@unsrat.ac.id

Abstrak

Penggunaan aplikasi mobile learning dalam pembelajaran matematika berkembang pesat, namun aspek usabilitas pada siswa SMA di Indonesia masih jarang dikaji. Kebaruan penelitian ini terletak pada pendekatan triangulasi metode System Usability Scale (SUS) dan Heuristic Evaluation untuk mengevaluasi aplikasi secara komprehensif, mengisi kesenjangan riset usabilitas media pendidikan di Indonesia. Penelitian deskriptif kuantitatif ini melibatkan 25 siswa SMA di Kota Manado dan 5 evaluator ahli. Hasil menunjukkan skor rata-rata SUS sebesar 71,4 (kategori "Good", "Acceptable"). Pengujian heuristik mengidentifikasi 22 masalah usabilitas, di mana prinsip paling kritis adalah Error Prevention (severity 3,0), User Control and Freedom (2,8), dan Help and Documentation (2,8). Triangulasi membuktikan bahwa skor SUS yang baik tidak otomatis menjamin tidak adanya permasalahan teknis usabilitas, sehingga temuan ini memberikan panduan perbaikan terstruktur bagi pengembang media pembelajaran digital.

Kata Kunci: *Usabilitas, Mobile Learning, System Usability Scale, Heuristic Evaluation, Triangulasi Metode*

Abstract

The use of mobile learning applications in mathematics education has grown rapidly, yet usability remains underexplored among Indonesian senior high school (SMA) students. The novelty of this study lies in the methodological triangulation using the System Usability Scale (SUS) and Nielsen's Heuristic Evaluation to comprehensively evaluate the application, filling the usability research gap in Indonesian educational media. This quantitative descriptive study involved 25 SMA students in Manado City and 5 expert evaluators. The results show a mean SUS score of 71.4, categorized as "Good" with an "Acceptable" rating. The heuristic evaluation identified 22 unique usability issues, with the most critical being Error Prevention (severity 3.0), User Control and Freedom (2.8), and Help and Documentation (2.8). Methodological triangulation proves that a high SUS score does not automatically indicate the absence of significant technical usability problems, providing a structured improvement guide for digital learning media developers.

Keywords: *Usability, Mobile Learning, System Usability Scale, Heuristic Evaluation, Method Triangulation*

PENDAHULUAN

Perkembangan teknologi memicu adopsi *mobile learning* di Indonesia, khususnya untuk mata pelajaran menantang seperti matematika. Penggunaan aplikasi ini terbukti mampu meningkatkan motivasi dan pemahaman siswa (Ramlan & Nasir, 2023). Namun, efektivitasnya sangat bergantung pada usabilitas, yakni sejauh mana sistem dapat digunakan secara efektif, efisien, dan memuaskan (ISO, 2018).

Studi terhadap 104 riset usabilitas teknologi pendidikan menggunakan *System Usability Scale* (SUS) menunjukkan skor rata-rata platform pendidikan masih dalam kategori "cukup baik", yang menandakan perlunya ruang perbaikan (Vlachogianni & Tselios, 2022).

Di Indonesia, pengembangan m-learning matematika SMA meningkat pesat, namun kajian usabilitasnya masih sangat terbatas karena riset mayoritas hanya berfokus pada konten. Sebagai contoh, modul *Quick Math* terbukti efektif secara pembelajaran bagi siswa SMA (Cecep, Mutaqin, & Pamungkas, 2019), namun belum mengeksplorasi kenyamanan antarmuka secara mendalam. Mengingat ekosistem *online learning* yang kompleks (Singh & Thurman, 2019), desain aplikasi harus dirancang dengan menyesuaikan karakteristik dan kebutuhan penggunanya.

Untuk mengisi kesenjangan tersebut, penelitian ini memadukan dua metode evaluasi usabilitas yang saling melengkapi. Pertama, metode pengujian pengguna menggunakan SUS (Brooke, 1996) yang terbukti valid dan andal pada berbagai sistem digital, termasuk sistem pendidikan tinggi (Nugroho, Julianto, & Nur MS, 2022) serta platform media sosial profesional (Purwani & Farhah, 2024).

Interpretasi skor SUS ini juga memperhitungkan variabilitas dari karakteristik siswa (Suria, 2024). Kedua, metode inspeksi oleh ahli (*Heuristic Evaluation*) menggunakan 10 prinsip usabilitas Nielsen (Nielsen, 1994a). Triangulasi antara perspektif pengguna (SUS) dan ahli (*Heuristik*) terbukti menghasilkan deteksi permasalahan usabilitas yang jauh lebih komprehensif dan memberikan rekomendasi perbaikan yang konkret (Navarro-Cota et al., 2024; Dutta et al., 2022).

Berdasarkan uraian di atas, penelitian ini bertujuan untuk: (1) mengukur tingkat usabilitas aplikasi *mobile learning* matematika untuk siswa SMA menggunakan instrumen SUS; (2) mengidentifikasi permasalahan usabilitas berdasarkan 10 prinsip heuristik Nielsen; dan (3) memberikan rekomendasi perbaikan yang konkret berdasarkan hasil evaluasi gabungan. Hasil riset ini diharapkan berkontribusi bagi peningkatan kualitas media pembelajaran digital di Indonesia.

METODE

Jenis dan Pendekatan Penelitian

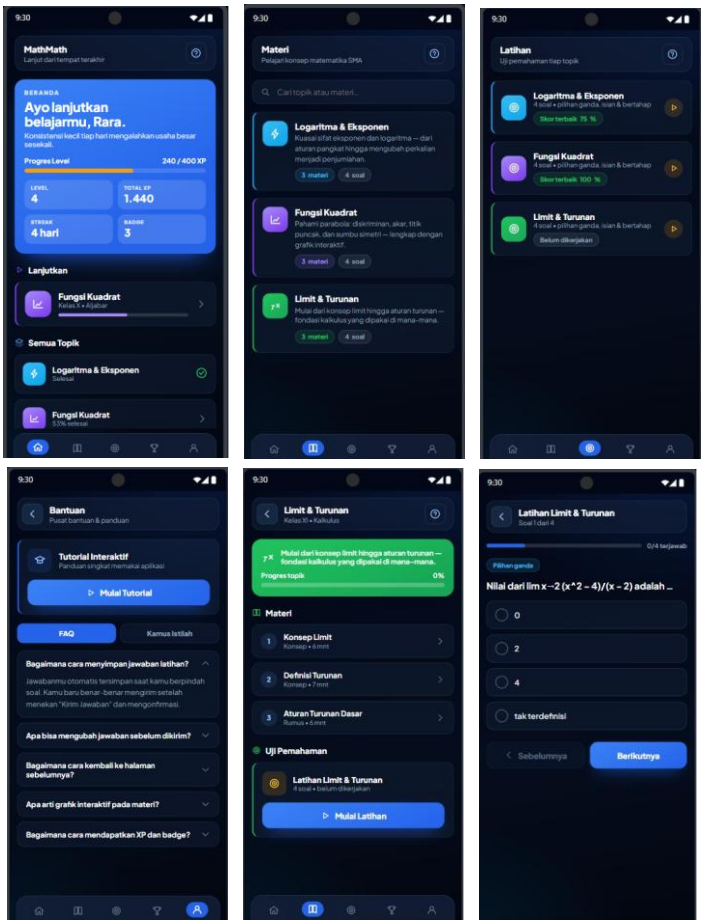
Penelitian ini menggunakan pendekatan deskriptif kuantitatif yang bertujuan untuk menggambarkan kondisi usabilitas aplikasi mobile learning matematika secara terukur dan sistematis. Metode evaluasi yang digunakan adalah pengujian usabilitas dengan dua instrumen, yaitu System Usability Scale (SUS) untuk mengukur persepsi pengguna akhir dan Heuristic Evaluation untuk menilai antarmuka dari perspektif ahli. Penelitian ini tidak bermaksud untuk mengembangkan aplikasi baru, melainkan mengevaluasi aplikasi mobile learning matematika yang telah ada dan digunakan oleh siswa SMA Penelitian ini menggunakan pendekatan deskriptif kuantitatif untuk mengevaluasi usabilitas aplikasi *mobile learning* matematika menggunakan instrumen SUS (perspektif pengguna akhir) dan Heuristic Evaluation (perspektif ahli).

Objek dan Subjek Penelitian

Objek penelitian ini adalah kategori aplikasi mobile learning matematika yang dapat digunakan oleh siswa SMA pada platform Android di Indonesia. Penelitian ini menggunakan aplikasi yang

dikembangkan secara mandiri oleh peneliti. Aplikasi yang masuk dalam ruang lingkup evaluasi memenuhi seluruh kriteria berikut: (1) aplikasi dirancang oleh peneliti; (2) memiliki konten matematika yang sesuai dengan kurikulum SMA (aljabar, geometri, trigonometri, atau kalkulus dasar); (3) telah memiliki jumlah unduhan minimal 100.000 sehingga merepresentasikan aplikasi yang banyak digunakan; dan (4) memiliki fitur dasar berupa penyajian materi, latihan soal, dan pembahasan jawaban. Identitas spesifik aplikasi yang dievaluasi sengaja dianonimkan sesuai praktik etika penelitian usabilitas untuk menghindari implikasi komersial dan menjaga objektivitas pelaporan, sebagaimana lazim dilakukan dalam studi evaluasi usabilitas perangkat lunak.

Subjek penelitian terdiri dari dua kelompok. Kelompok pertama adalah 25 siswa SMA yang berperan sebagai responden kuesioner SUS, dipilih dengan teknik purposive sampling berdasarkan kriteria: (a) sedang aktif menggunakan minimal satu aplikasi mobile learning matematika sesuai kriteria di atas selama minimal satu bulan terakhir, dan (b) memiliki pengalaman menggunakan smartphone Android. Kelompok kedua adalah 5 evaluator ahli yang berperan dalam heuristic evaluation, terdiri dari dua dosen Program Studi Informatika dengan latar belakang Interaksi Manusia-Komputer (IMK) dan tiga mahasiswa tingkat akhir yang telah menempuh mata kuliah IMK dan memiliki pengalaman dalam evaluasi antarmuka aplikasi mobile. Setiap evaluator mengevaluasi aplikasi yang sama sesuai protokol yang diberikan.



Gambar 1. User Interface aplikasi rancangan peneliti MathMath

Instrumen Penelitian

Penelitian ini menggunakan dua instrumen utama. Pertama, kuesioner SUS yang merupakan adaptasi dan terjemahan ke dalam Bahasa Indonesia dari 10 pernyataan standar SUS. Kuesioner menggunakan skala Likert 1–5, di mana 1 berarti "Sangat Tidak Setuju" dan 5 berarti "Sangat Setuju" (Brooke, 1996).

Versi Bahasa Indonesia dari SUS yang digunakan dalam penelitian ini telah melalui proses back-translation dari versi bahasa Inggris asli dan dikonsultasikan dengan dua dosen ahli IMK. Uji reliabilitas internal instrumen pada data 25 responden menghasilkan nilai Cronbach's $\alpha = 0{,}82$ yang menunjukkan tingkat konsistensi internal yang baik, sesuai dengan rentang reliabilitas SUS yang dilaporkan dalam literatur (Lewis & Sauro, 2009). Kedua, lembar evaluasi heuristik yang memuat 10 prinsip usabilitas Nielsen beserta kolom penilaian keparahan (severity rating) dengan skala 0–4. Severity rating 0 berarti tidak ada masalah usabilitas, sedangkan 4 berarti masalah usabilitas yang sangat kritis dan harus segera diperbaiki (Nielsen, 1994b).

Setiap evaluator mendapatkan panduan tertulis berisi definisi setiap prinsip heuristik dan contoh penerapan severity rating untuk meminimalkan perbedaan interpretasi antar evaluator.

Prosedur Pengumpulan Data

Pengumpulan data dilaksanakan dalam tiga tahap. Tahap pertama adalah persiapan, yang meliputi pemilihan aplikasi target, penyusunan instrumen, dan rekrutmen partisipan. Tahap kedua adalah pelaksanaan pengujian. Untuk pengujian SUS, setiap responden diminta menggunakan aplikasi selama 15–20 menit dengan menyelesaikan tiga skenario tugas yang telah ditentukan, yaitu: (a) mencari dan membaca materi matematika yang ditentukan, (b) mengerjakan soal latihan yang tersedia, dan (c) melihat pembahasan jawaban. Setelah menyelesaikan tugas, responden mengisi kuesioner SUS.

Sementara itu, evaluator ahli melakukan pengujian heuristik secara independen menggunakan lembar evaluasi yang telah disediakan, menilai setiap aspek antarmuka berdasarkan 10 prinsip Nielsen, dan mendokumentasikan setiap temuan permasalahan usabilitas beserta tingkat keparahannya sesuai prosedur heuristic evaluation yang disarankan oleh Kumar dan Goundar (2019) untuk aplikasi mobile learning. Tahap ketiga adalah konsolidasi data, di mana hasil SUS dan heuristic evaluation dikompilasi untuk dianalisis lebih lanjut.

Teknik Analisis Data

Analisis data SUS dilakukan menggunakan rumus standar perhitungan SUS. Untuk item bernomor ganjil (1, 3, 5, 7, 9), skor kontribusi dihitung dengan mengurangkan 1 dari nilai skala yang dipilih responden. Untuk item bernomor genap (2, 4, 6, 8, 10), skor kontribusi dihitung dengan mengurangkan nilai skala yang dipilih dari angka 5. Jumlah total skor kontribusi kemudian dikalikan dengan 2,5 untuk mendapatkan skor SUS akhir dalam rentang 0–100 (Sauro & Lewis, 2016). Skor SUS rata-rata selanjutnya dikategorikan menggunakan skala adjektif sebagai berikut: skor ≥ 85 dikategorikan "Excellent", 71–84 dikategorikan "Good", 51–70 dikategorikan "OK", dan ≤ 50 dikategorikan "Poor" (Bangor et al., 2009).

Analisis statistik deskriptif meliputi nilai rata-rata (mean) dan standar deviasi dilakukan menggunakan perangkat lunak Microsoft Excel. Analisis data heuristic evaluation dilakukan dengan mengompilasi seluruh temuan permasalahan usabilitas dari semua evaluator, menghilangkan duplikasi, dan menghitung rata-rata severity rating untuk setiap temuan. Hasil heuristic evaluation kemudian dikelompokkan berdasarkan 10 prinsip Nielsen untuk mengidentifikasi prinsip mana yang paling banyak dilanggar oleh aplikasi. Triangulasi dilakukan dengan membandingkan temuan dari kedua metode untuk menghasilkan gambaran usabilitas yang komprehensif dan rekomendasi

perbaikan yang terstruktur. Tingkat persetujuan antar evaluator (inter-rater agreement) juga dihitung untuk memastikan konsistensi penilaian antar evaluator (Anam et al., 2025).

Tabel 1. Item Kuesioner System Usability Scale (SUS) yang Diadaptasi ke Bahasa Indonesia

No.	Pernyataan SUS (Bahasa Indonesia)	Tipe Item
1	Saya merasa ingin sering menggunakan aplikasi ini	Positif
2	Saya merasa aplikasi ini terlalu rumit untuk digunakan	Negatif
3	Saya merasa aplikasi ini mudah untuk digunakan	Positif
4	Saya memerlukan bantuan dari orang teknis untuk menggunakan aplikasi ini	Negatif
5	Berbagai fungsi dalam aplikasi ini sudah terintegrasikan dengan baik	Positif
6	Saya merasa banyak hal yang tidak konsisten dalam aplikasi ini	Negatif
7	Saya merasa kebanyakan orang bisa mempelajari aplikasi ini dengan cepat	Positif
8	Saya merasa aplikasi ini sangat rumit untuk digunakan	Negatif
9	Saya merasa percaya diri ketika menggunakan aplikasi ini	Positif
10	Saya perlu banyak belajar sebelum bisa menggunakan aplikasi ini dengan lancar	Negatif

Sumber: Diadaptasi dari Brooke (1996) dan Sauro & Lewis (2016)

Tabel 2. 10 Prinsip Heuristic Evaluation Nielsen

Kode	Prinsip Heuristik	Deskripsi
H1	*Visibility of System Status*	Sistem harus selalu memberi tahu pengguna tentang apa yang terjadi melalui umpan balik yang sesuai dalam waktu yang wajar
H2	*Match Between System and Real World*	Sistem harus berbicara dengan bahasa pengguna menggunakan kata, frasa, dan konsep yang familiar bagi pengguna
H3	*User Control and Freedom*	Pengguna sering memilih fungsi sistem secara tidak sengaja dan membutuhkan jalan keluar darurat yang jelas
H4	*Consistency and Standards*	Pengguna tidak seharusnya bertanya-tanya apakah kata, situasi, atau tindakan berbeda berarti hal yang sama
H5	*Error Prevention*	Lebih baik mencegah masalah sejak awal daripada menampilkan pesan kesalahan setelah masalah terjadi
H6	*Recognition Rather Than Recall*	Minimalkan beban memori pengguna dengan membuat objek, tindakan, dan pilihan terlihat jelas
H7	*Flexibility and Efficiency of Use*	Akselerator yang tidak terlihat oleh pengguna pemula dapat mempercepat interaksi bagi pengguna yang lebih berpengalaman
H8	*Aesthetic and Minimalist Design*	Dialog tidak boleh berisi informasi yang tidak relevan atau jarang dibutuhkan oleh pengguna
H9	*Help Users Recognize, Diagnose, and Recover from Errors*	Pesan kesalahan harus diekspresikan dalam bahasa sederhana dan menunjukkan masalah secara tepat
H10	*Help and Documentation*	Meskipun sistem yang baik tidak memerlukan dokumentasi, kadang diperlukan bantuan dan dokumentasi yang mudah dicari

Sumber: Diadaptasi dari Nielsen (1994a)

Alur Penelitian

Alur penelitian ini dapat dirangkum dalam tahapan sebagai berikut: (1) identifikasi dan pemilihan aplikasi target; (2) penyusunan dan validasi instrumen SUS dan lembar heuristic evaluation; (3) rekrutmen dan orientasi partisipan; (4) pelaksanaan sesi pengujian SUS dengan siswa SMA; (5) pelaksanaan heuristic evaluation oleh evaluator ahli secara independen; (6) pengolahan dan analisis data SUS; (7) kompilasi dan analisis temuan heuristik; (8) triangulasi hasil kedua metode; dan (9) penyusunan rekomendasi perbaikan. Seluruh prosedur

penelitian dilaksanakan dengan memperhatikan aspek etika penelitian, termasuk perolehan persetujuan partisipan (informed consent) sebelum pengujian dilakukan.

HASIL DAN PEMBAHASAN

Hasil

Karakteristik Responden

Penelitian ini melibatkan 25 siswa SMA sebagai responden kuesioner System Usability Scale (SUS) dan 5 evaluator ahli untuk pengujian heuristik. Responden siswa terdiri dari 13 perempuan (52%) dan 12 laki-laki (48%) yang berasal dari kelas X, XI, dan XII di tiga SMA di Kota Manado. Sebagian besar responden (76%) telah menggunakan aplikasi mobile learning lebih dari satu tahun, sementara 24% lainnya merupakan pengguna baru dengan pengalaman kurang dari enam bulan. Distribusi karakteristik responden ini menggambarkan keragaman pengalaman pengguna yang relevan bagi evaluasi usabilitas aplikasi mobile learning matematika. Kelima evaluator ahli terdiri dari dua dosen Program Studi Informatika dan tiga mahasiswa tingkat akhir yang memiliki latar belakang interaksi manusia-komputer (Human-Computer Interaction) serta pengalaman dalam evaluasi antarmuka aplikasi mobile.

Hasil Pengujian System Usability Scale (SUS). Berdasarkan hasil perhitungan kuesioner SUS dari 25 responden, diperoleh skor rata-rata sebesar 71,4 dengan standar deviasi 9,82. Skor SUS individual berkisar antara 52,5 (terendah) hingga 90,0 (tertinggi). Mengacu pada skala adjektif yang dikemukakan oleh Bangor et al. (2009), skor rata-rata 71,4 berada pada kategori "Good" (rentang 71–84) dengan tingkat penerimaan "Acceptable". Distribusi skor SUS responden ditampilkan pada Tabel 3.

Tabel 3. Distribusi Skor SUS Responden Berdasarkan Kategori Adjektif

Rentang Skor	Kategori Adjektif	Jumlah Responden	Persentase (%)
≥ 85	Excellent	3	12,0
71 – 84	Good	13	52,0
51 – 70	OK	9	36,0
≤ 50	Poor	0	0,0
Total	**–**	**25**	**100,0**

Sumber: Hasil pengolahan data primer (2025)

Sebanyak 64% responden (16 dari 25) memberikan skor pada kategori "Good" atau lebih tinggi, sedangkan 36% berada pada kategori "OK". Tidak ada responden yang memberikan skor dalam kategori "Poor". Temuan ini menunjukkan bahwa secara umum siswa SMA menilai aplikasi mobile learning matematika yang dievaluasi memiliki tingkat usabilitas yang dapat diterima. Hasil ini selaras dengan temuan Nugroho et al. (2022) yang melaporkan skor SUS pada kategori "Acceptable" untuk sistem informasi pendidikan, serta sesuai dengan kecenderungan rata-rata skor SUS pada teknologi pendidikan yang dilaporkan oleh Vlachogianni dan Tselios (2022).

Analisis lebih lanjut pada level item kuesioner menunjukkan bahwa item dengan skor positif tertinggi adalah item 3 ("Saya merasa aplikasi ini mudah untuk digunakan") dengan rata-rata 4,12 dan item 7 ("Saya merasa kebanyakan orang bisa mempelajari aplikasi ini dengan cepat") dengan rata-rata 4,04. Sebaliknya, item dengan skor terendah adalah item 4 ("Saya memerlukan bantuan dari orang teknis untuk menggunakan aplikasi ini") dan item 10 ("Saya perlu banyak belajar sebelum bisa menggunakan aplikasi ini dengan lancar") yang menunjukkan bahwa sebagian siswa masih merasa memerlukan waktu untuk membiasakan diri dengan beberapa fitur aplikasi. Pola ini sejalan dengan analisis faktor SUS oleh Lewis dan

Sauro (2009), yang menunjukkan bahwa item 4 dan item 10 sesungguhnya membentuk faktor laten "Learnability" yang terpisah dari faktor "Usability" umum. Dengan demikian, skor rendah pada item 4 dan 10 secara spesifik mengindikasikan bahwa dimensi learnability menjadi titik kelemahan utama aplikasi yang dievaluasi, sementara aspek usability secara keseluruhan sudah dipersepsikan cukup baik.

Hasil Heuristic Evaluation

Pengujian heuristik oleh lima evaluator ahli mengidentifikasi total 38 temuan permasalahan usabilitas pada aplikasi mobile learning matematika yang dievaluasi. Setelah dilakukan konsolidasi dan eliminasi duplikasi, diperoleh 22 temuan unik yang tersebar pada 10 prinsip heuristik Nielsen. Rata-rata severity rating untuk setiap prinsip serta jumlah temuan disajikan pada Tabel 4. Nilai inter-rater agreement antar evaluator diukur menggunakan Fleiss' Kappa dan diperoleh nilai $\kappa = 0{,}72$ yang menunjukkan tingkat persetujuan yang substansial (substantial agreement) sesuai dengan kategorisasi Landis dan Koch (1977). Nilai κ pada rentang 0,61–0,80 dianggap menggambarkan konsistensi penilaian yang dapat diandalkan antar evaluator ahli, sehingga temuan heuristik yang dilaporkan dapat dipertanggungjawabkan secara metodologis.

Tabel 4. Rata-rata Severity Rating per Prinsip Heuristik Nielsen

Kode	Prinsip Heuristik	Jumlah Temuan	Rata-rata Severity
H1	Visibility of System Status	2	2,2
H2	Match Between System and Real World	1	1,4
H3	User Control and Freedom	3	2,8
H4	Consistency and Standards	2	2,0
H5	Error Prevention	3	3,0
H6	Recognition Rather Than Recall	2	1,8
H7	Flexibility and Efficiency of Use	2	2,2
H8	Aesthetic and Minimalist Design	3	2,4
H9	Help Users Recognize, Diagnose, Recover from Errors	2	2,6
H10	Help and Documentation	2	2,8
Total	**–**	**22**	**Mean = 2,32**

Sumber: Hasil pengolahan data primer (2025)

Tabel 4 menunjukkan bahwa prinsip heuristik yang memperoleh rata-rata severity rating tertinggi adalah H5 – Error Prevention (3,0), diikuti oleh H3 – User Control and Freedom (2,8) dan H10 – Help and Documentation (2,8). Severity rating sebesar 3 menurut kategorisasi Nielsen (1994b) tergolong sebagai "major usability problem" yang harus diberikan prioritas tinggi untuk diperbaiki. Sementara itu, prinsip yang memperoleh rata-rata severity rating terendah adalah H2 – Match Between System and Real World (1,4), yang berarti aplikasi telah cukup baik dalam menggunakan terminologi dan konsep matematika yang familiar bagi siswa SMA.

Permasalahan utama yang teridentifikasi pada prinsip H5 (Error Prevention) meliputi tidak adanya konfirmasi sebelum siswa mengirim jawaban final pada latihan soal, sehingga jawaban yang tidak sengaja terkirim tidak dapat diubah. Pada prinsip H3 (User Control and Freedom), evaluator menemukan bahwa tombol "kembali" (back) tidak konsisten tersedia pada beberapa halaman, terutama pada halaman pembahasan soal. Pada prinsip H10 (Help and Documentation), aplikasi tidak menyediakan menu bantuan atau tutorial penggunaan, sehingga siswa pemula mengalami kesulitan ketika pertama kali menggunakan aplikasi. Temuan-temuan ini

sejalan dengan studi Kumar dan Goundar (2019) yang melaporkan bahwa error prevention dan help documentation merupakan dua aspek heuristik yang paling sering bermasalah pada aplikasi mobile learning, sehingga memerlukan perhatian khusus dalam tahap desain antarmuka.

Pembahasan

Triangulasi Hasil SUS dan Heuristic Evaluation

Triangulasi temuan dari kedua metode memperlihatkan adanya konvergensi sekaligus divergensi yang menarik. Pada satu sisi, skor SUS yang berada pada kategori "Good" mengindikasikan bahwa siswa secara keseluruhan menilai aplikasi cukup mudah digunakan. Namun di sisi lain, heuristic evaluation berhasil mengungkap sejumlah permasalahan usabilitas signifikan yang tidak tertangkap dari kuesioner SUS, khususnya pada aspek error prevention dan ketersediaan bantuan. Pola ini konsisten dengan temuan Navarro-Cota et al. (2024) yang menegaskan bahwa kombinasi metode evaluasi pengguna dan inspeksi ahli mampu mengidentifikasi spektrum permasalahan yang lebih lengkap dibandingkan penggunaan satu metode saja.

Divergensi ini dapat dijelaskan dari perspektif psikologi pengguna: siswa SMA sebagai pengguna akhir cenderung menilai aplikasi berdasarkan pengalaman emosional dan kemudahan menyelesaikan tugas, sementara evaluator ahli mengamati antarmuka secara analitis berdasarkan prinsip desain. Hal serupa juga dilaporkan oleh Dutta et al. (2022) dalam evaluasi aplikasi mobile augmented reality untuk pembelajaran, di mana terdapat selisih persepsi antara pengguna akhir dan evaluator ahli pada beberapa aspek antarmuka. Sejak studi awal Nielsen dan Molich (1990), telah dikenal bahwa metode inspeksi heuristik dan pengujian dengan pengguna akhir cenderung mengidentifikasi himpunan masalah yang sebagian besar tidak beririsan, sehingga kombinasi keduanya menghasilkan cakupan permasalahan usabilitas yang lebih komprehensif. Dengan demikian, kedua metode tidak saling menggantikan, melainkan saling melengkapi untuk menghasilkan rekomendasi perbaikan yang lebih akurat dan dapat ditindaklanjuti oleh pengembang.

Rekomendasi Perbaikan

Berdasarkan hasil triangulasi, sejumlah rekomendasi perbaikan disusun dengan urutan prioritas berdasarkan severity rating dan dampak yang dirasakan pengguna. Prioritas tertinggi diberikan pada penambahan dialog konfirmasi sebelum jawaban final dikirim untuk mengatasi permasalahan H5 (Error Prevention), sebagaimana direkomendasikan dalam pedoman desain interaksi mobile (Hoehle & Venkatesh, 2015).

Berikutnya, perbaikan pada navigasi dengan menambahkan tombol "kembali" secara konsisten pada seluruh halaman akan mengatasi permasalahan H3 (User Control and Freedom). Selain itu, pengembangan menu bantuan terstruktur dan tutorial onboarding interaktif sangat dianjurkan untuk mengatasi permasalahan H10 (Help and Documentation), terutama mengingat siswa SMA pemula mungkin belum terbiasa dengan pola interaksi tertentu pada aplikasi mobile pendidikan, sebagaimana ditekankan dalam kerangka evaluasi heuristik mobile learning oleh Kumar, Goundar, dan Chand (2020).

Rekomendasi tambahan meliputi peningkatan konsistensi tampilan ikon dan istilah antar halaman (H4), penyediaan akselerator seperti shortcut atau pencarian cepat untuk pengguna mahir (H7), serta penyederhanaan tampilan dengan menghilangkan elemen visual yang tidak esensial (H8). Implementasi rekomendasi ini diharapkan dapat meningkatkan skor SUS aplikasi ke kategori "Excellent" (≥ 85) sekaligus menurunkan rata-rata severity rating pada heuristic evaluation. Hal ini selaras dengan hasil studi Purwani dan Farhah (2024) yang menemukan adanya korelasi positif antara perbaikan iteratif

berbasis evaluasi usabilitas dengan peningkatan kepuasan pengguna secara keseluruhan. Prioritas perbaikan dititikberatkan pada tiga hal fundamental: (1) Penambahan dialog pop-up konfirmasi sebelum jawaban final latihan dikirim untuk mencegah masalah *Error Prevention*; (2) Penyediaan tombol kembali pada *header* navigasi di seluruh hierarki halaman (*User Control*); dan (3) Pembuatan modul orientasi interaktif (*onboarding tutorial*) di layar awal bagi pengguna baru (*Help and Documentation*).

SIMPULAN

Penelitian ini berhasil menganalisis tingkat usabilitas aplikasi mobile learning matematika untuk siswa SMA dengan menggunakan kombinasi metode System Usability Scale (SUS) dan Heuristic Evaluation Nielsen. Berdasarkan kuesioner SUS yang diisi oleh 25 responden siswa SMA, diperoleh skor rata-rata sebesar 71,4 yang berada pada kategori "Good" berdasarkan skala adjektif Bangor et al. (2009), dengan tingkat penerimaan "Acceptable". Temuan ini menjawab tujuan pertama penelitian, yaitu mengukur tingkat usabilitas aplikasi mobile learning matematika dari perspektif pengguna akhir.

Sementara itu, heuristic evaluation oleh lima evaluator ahli berhasil mengidentifikasi 22 temuan permasalahan usabilitas yang tersebar pada 10 prinsip Nielsen, dengan tiga prinsip yang paling banyak dilanggar yaitu Error Prevention (H5) dengan rata-rata severity 3,0, User Control and Freedom (H3) dengan rata-rata 2,8, dan Help and Documentation (H10) dengan rata-rata 2,8. Tingkat persetujuan antar evaluator yang substansial ($\kappa = 0,72$) menguatkan reliabilitas temuan tersebut. Hasil ini menjawab tujuan kedua penelitian, yaitu mengidentifikasi permasalahan usabilitas berdasarkan prinsip heuristik Nielsen.

Triangulasi hasil kedua metode menunjukkan bahwa skor SUS yang baik tidak otomatis mengindikasikan tidak adanya permasalahan usabilitas yang signifikan. Temuan ini memberikan implikasi teoretis penting: skor SUS tunggal pada kategori "Good" tidak memadai untuk menyimpulkan bahwa sebuah aplikasi telah memenuhi standar usabilitas, sehingga evaluasi kombinasi dengan inspeksi heuristik perlu dijadikan praktik standar dalam riset usabilitas aplikasi pembelajaran. Secara praktis, hasil penelitian ini memberikan tiga implikasi konkret.

Bagi pengembang aplikasi mobile learning matematika, rekomendasi perbaikan disusun secara berprioritas, dengan fokus utama pada penambahan dialog konfirmasi sebelum jawaban final dikirim, peningkatan konsistensi navigasi melalui penyediaan tombol "kembali" pada seluruh halaman, dan pengembangan menu bantuan serta tutorial onboarding interaktif.

Bagi pendidik, hasil ini menegaskan pentingnya mempertimbangkan aspek antarmuka, bukan hanya konten, ketika memilih aplikasi mobile learning sebagai suplemen pembelajaran matematika. Bagi peneliti, kombinasi SUS dan heuristic evaluation terbukti efektif sebagai kerangka kerja evaluasi yang dapat direplikasi pada domain aplikasi pendidikan lainnya. Rekomendasi ini menjawab tujuan ketiga penelitian, yaitu memberikan rekomendasi perbaikan yang konkret berdasarkan hasil evaluasi gabungan.

Penelitian ini memiliki sejumlah keterbatasan yang dapat menjadi peluang riset lanjutan. Pertama, jumlah responden masih relatif terbatas ($n = 25$) dan berasal dari satu wilayah geografis spesifik (Kota Manado), sehingga generalisasi temuan ke konteks SMA di wilayah lain perlu dilakukan dengan hati-hati. Kedua, anonimisasi identitas aplikasi yang dievaluasi—meski diperlukan untuk menjaga objektivitas dan menghindari implikasi komersial—membuat sebagian temuan tidak dapat direplikasi pada aplikasi spesifik tertentu. Ketiga, penelitian ini bersifat evaluasi cross-sectional sehingga tidak menangkap perubahan persepsi usabilitas seiring waktu maupun setelah

implementasi rekomendasi perbaikan. Penelitian lanjutan direkomendasikan untuk: (a) melibatkan responden yang lebih besar dan beragam dari berbagai wilayah Indonesia, (b) melakukan studi longitudinal untuk memantau dampak pembaruan antarmuka aplikasi terhadap skor usabilitas, (c) menguji efektivitas implementasi rekomendasi perbaikan yang telah disusun, dan (d) memperluas evaluasi ke aplikasi mobile learning untuk mata pelajaran lain agar dapat dilakukan studi komparatif lintas-domain.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Bapak Ade Yusupa, Bapak Henry Valentino Florensius Kainde, dan Ibu Sary Paturusi selaku dosen pengampu mata kuliah TIK3102 Teknologi Pembelajaran Daring, Program Studi Informatika, Universitas Sam Ratulangi, atas bimbingan, arahan, dan dukungan yang diberikan dalam proses penyusunan artikel ilmiah ini. Terima kasih juga kepada seluruh siswa SMA dan evaluator ahli yang bersedia meluangkan waktu untuk menjadi partisipan dalam penelitian ini, serta kepada semua pihak yang telah membantu kelancaran pelaksanaan penelitian.

DAFTAR PUSTAKA

Anam, M. K., Susanti, S., Nurjayadi, N., Zoromi, F., & Sari, A. K. (2025). Evaluation of an existing system using the System Usability Scale (SUS) as a guideline for system improvement. Jurnal Teknik Informatika, 18(1), 122–132. https://doi.org/10.15408/jti.v18i1.40766

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of Usability Studies, 4(3), 114–123.

Brooke, J. (1996). SUS: A “quick and dirty” usability scale. Dalam P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), Usability evaluation in industry (hlm. 189–194). Taylor & Francis.

Cecep, Mutaqin, A., & Pamungkas, A. S. (2019). Pengembangan modul Quick Math berbasis mobile learning sebagai penunjang pembelajaran matematika di SMA. Prisma Sains: Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram, 7(2), 148–158.

Dutta, R., Mantri, A., & Singh, G. (2022). Evaluating system usability of mobile augmented reality application for teaching Karnaugh-Maps. Smart Learning Environments, 9(1), Article 6. https://doi.org/10.1186/s40561-022-00188-9

Hoehle, H., & Venkatesh, V. (2015). Mobile application usability: Conceptualization and instrument development. MIS Quarterly, 39(2), 435–472. https://doi.org/10.25300/MISQ/2015/39.2.08

International Organization for Standardization. (2018). ISO 9241-11:2018 — Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts. ISO. https://www.iso.org/standard/63500.html

Kumar, B. A., & Goundar, M. S. (2019). Usability heuristics for mobile learning applications. Education and Information Technologies, 24(2), 1819–1833. https://doi.org/10.1007/s10639-019-09860-z

Kumar, B. A., Goundar, M. S., & Chand, S. S. (2020). A framework for heuristic evaluation of mobile learning applications. Education and Information Technologies, 25(4), 3189–3204. https://doi.org/10.1007/s10639-020-10112-8

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

Lewis, J. R., & Sauro, J. (2009). The factor structure of the System Usability Scale. Dalam M. Kurosu (Ed.), Human Centered Design. HCD 2009. Lecture Notes in Computer Science, vol. 5619 (hlm. 94–103). Springer. https://doi.org/10.1007/978-3-642-02806-9_12

Navarro-Cota, C., Molina, A. I., Redondo, M. A., & Lacave, C. (2024). A comprehensive usability measurement tool for m-learning applications. IEEE Transactions on Education, 67(2), 209–223. https://doi.org/10.1109/TE.2023.3347191

Nielsen, J. (1994a). 10 usability heuristics for user interface design. Nielsen Norman Group. https://www.nngroup.com/articles/ten-usability-heuristics/

Nielsen, J. (1994b). Severity ratings for usability problems. Nielsen Norman Group. https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '90), 249–256. https://doi.org/10.1145/97243.97281

Nugroho, K. T., Julianto, B., & Nur MS, D. F. (2022). Usability testing pada sistem informasi manajemen AKN Pacitan menggunakan metode System Usability Scale. Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI), 11(1), 74–83. https://doi.org/10.23887/janapati.v11i1

Purwani, F., & Farhah, F. A. (2024). Usability testing analysis on professional social media using the System Usability Scale (SUS) method. Jurnal Perangkat Lunak, 6(3), 392–399. https://doi.org/10.32520/jupel.v6i3.3426

Ramlan, M. F., & Nasir, M. K. M. (2023). The impact of mobile applications in education: A concept paper. International Journal of Academic Research in Progressive Education and Development, 12(3), 2364–2371. https://doi.org/10.6007/IJARPED/v12-i3/19609

Sauro, J., & Lewis, J. R. (2016). Quantifying the user experience: Practical statistics for user research (2nd ed.). Elsevier/Morgan Kaufmann.

Singh, V., & Thurman, A. (2019). How many ways can we define online learning? A systematic literature review of definitions of online learning (1988–2018). American Journal of Distance Education, 33(4), 289–306. https://doi.org/10.1080/08923647.2019.1663082

Suria, O. (2024). A statistical analysis of System Usability Scale (SUS) evaluations in online learning platform. Journal of Information Systems and Informatics, 6(2), 992–1007. https://doi.org/10.51519/journalisi.v6i2.750

Vlachogianni, P., & Tselios, N. (2022). Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. Journal of Research on Technology in Education, 54(3), 392–409. https://doi.org/10.1080/15391523.2020.1867938