Pengembangan Aplikasi PAKSALEH untuk Asesmen dan Pengembangan Kompetensi Adaptif pada K3 Maritim

Muhammad Hidayat Lammi[1*], Muh. Syilfa Nooviar[2]
[1,2] Fakultas Ilmu Sosial dan Hukum, Administrasi Publik, Universitas Negeri Makassar
[1*]muhammad.hidayat@unm.ac.id ; [2]syilfa.nooviar@unm.ac.id
*Penulis korespondensi:
(**Muhammad Hidayat Lammi,**muhammad.hidayat@unm.ac.id)

ABSTRAK

Asesmen digital pada pendidikan dan pelatihan keselamatan dan kesehatan kerja (K3) maritim perlu menghasilkan informasi yang dapat digunakan untuk menentukan tindak lanjut pengembangan kompetensi. Penelitian ini bertujuan mendeskripsikan pengembangan aplikasi PAKSALEH dan mengevaluasi kualitasnya berdasarkan persepsi pengguna pada enam karakteristik ISO 9126. Penelitian menggunakan pendekatan penelitian dan pengembangan dengan kerangka Analysis, Design, Development, Implementation, dan Evaluation (ADDIE). Rancangan produk menghubungkan asesmen, penghitungan skor, pemetaan kompetensi, identifikasi kesenjangan, rekomendasi pengembangan, asesmen ulang, dan laporan perkembangan. Evaluasi dianalisis dari 104 responden menggunakan 18 butir pernyataan berskala 1–5, dengan tiga butir untuk setiap karakteristik. Analisis deskriptif mencakup rata-rata butir, rata-rata karakteristik, simpangan baku skor komposit responden, dan distribusi jawaban. Seluruh 1.872 jawaban memiliki nilai yang lengkap dan berada pada rentang skala. Rata-rata keseluruhan mencapai 4,35 dengan simpangan baku 0,76. Nilai functionality adalah 4,41, reliability 4,20, usability 4,34, efficiency 4,36, maintainability 4,38, dan portability 4,42. Sebanyak 81,20% jawaban berada pada skor 4–5. Dengan kategori deskriptif berdasarkan rata-rata yang dibulatkan dua desimal, lima karakteristik termasuk sangat baik dan reliability termasuk baik. Portability merupakan karakteristik dengan rata-rata tertinggi, sedangkan reliability menjadi prioritas evaluasi lanjutan. Temuan memberikan dasar untuk penyempurnaan stabilitas sistem dan penguatan informasi tindak lanjut bagi pengguna. Kesimpulan terbatas pada penilaian pengguna terhadap kualitas aplikasi; efektivitas peningkatan kompetensi, ketepatan rekomendasi adaptif, dan kualitas teknis internal belum diuji melalui data ini. Pengujian berbasis tugas dan validasi ahli diperlukan untuk melengkapi bukti pengembangan produk.

Kata Kunci: PAKSALEH; K3 maritim; asesmen kompetensi; pengembangan aplikasi; ISO 9126

ABSTRACT

Digital assessment in maritime occupational safety and health education should provide information that supports subsequent competence development. This study describes the development of the PAKSALEH application and evaluates its perceived quality using six ISO 9126 characteristics. A research and development approach was organized around the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) framework. The product design connects assessment, scoring, competence mapping, gap identification, development recommendations, reassessment, and progress reporting. The evaluation dataset comprised 104 respondents who completed 18 statements on a five-point scale, with three statements representing each characteristic. Descriptive analysis included item means, characteristic means, standard deviations of respondent composite scores, and response frequencies. All 1,872 responses were complete and within the permitted scale. The overall mean was 4.35, with a standard deviation of 0.76. Characteristic means were 4.41 for functionality, 4.20 for reliability, 4.34 for usability, 4.36 for efficiency, 4.38 for maintainability, and 4.42 for portability. Responses scored four or five accounted for 81.20% of all answers. Under descriptive categories applied after rounding means to two decimal places, five characteristics were rated very good and reliability was rated good. Portability had the highest mean, while reliability indicated a priority for further evaluation. These findings support application refinement, particularly system stability and actionable feedback. The evidence concerns perceived application quality and does not establish competence gains, recommendation accuracy, or internal technical quality. Further task-based testing and expert validation are needed to strengthen the evidence for the product.

Keywords: PAKSALEH; maritime occupational safety; competence assessment; application development; ISO 9126

1. PENDAHULUAN

Keselamatan dan kesehatan kerja (K3) maritim memerlukan pembelajaran yang menghubungkan pengetahuan dengan kemampuan memahami risiko dan menentukan tindakan dalam situasi kerja. Teknologi digital menyediakan lingkungan latihan dan asesmen yang lebih beragam. Dewan et al. [1] memetakan penggunaan simulator imersif dan nonimersif dalam pendidikan maritim, termasuk pelatihan keselamatan. Namun, mutu representasi teknologi perlu dibedakan dari mutu hasil belajar. De Oliveira et al. [2] menunjukkan bahwa fidelitas simulator mencakup dimensi fisik dan fungsional, sementara bukti pengaruh setiap dimensinya terhadap kinerja peserta masih terbatas. Karimi et al. [3] kemudian mengulas sistem instruksional adaptif untuk pelatihan keselamatan maritim dan menyoroti kebutuhan pemodelan peserta serta keterbatasan data. Ketiga kajian tersebut menunjukkan relevansi pengembangan teknologi yang dapat menghubungkan pengalaman belajar dengan informasi tentang kebutuhan kompetensi peserta.

Tantangan asesmen digital bukan hanya memindahkan pertanyaan ke layar, melainkan menjadikan hasilnya berguna bagi proses belajar berikutnya. Jurāne-Brēmane [4] mengidentifikasi peluang dan persoalan asesmen digital, termasuk umpan balik, validitas, dan transparansi. Tinjauan Solis Trujillo et al. [5] juga menempatkan umpan balik yang spesifik dan penggunaan teknologi sebagai unsur penting dalam asesmen formatif. Dalam rancangan PAKSALEH, persoalan yang hendak ditangani adalah kebutuhan menghubungkan skor asesmen K3 maritim dengan pemetaan kompetensi dan rekomendasi pengembangan individual. Apabila hasil hanya disajikan sebagai skor agregat, peserta dapat mengetahui capaian umum tetapi belum memperoleh informasi yang cukup untuk menentukan bagian kompetensi yang perlu diperkuat. Kebutuhan ini menjadi dasar perancangan sistem, bukan klaim bahwa seluruh sistem asesmen yang telah ada hanya menyimpan nilai.

Penelitian sebelumnya memberikan landasan bagi personalisasi tindak lanjut asesmen. Bayly-Castaneda et al. [6] meninjau pengembangan jalur belajar personal berbantuan kecerdasan buatan dan menekankan pentingnya desain pedagogis. Ihichr et al. [7] mengulas teori, algoritma, dan teknik asesmen adaptif dalam lingkungan pembelajaran digital. Gligorea et al. [8] membahas pemanfaatan kecerdasan buatan dan pembelajaran mesin dalam *e-learning* adaptif, termasuk manfaat serta tantangan integrasinya. Zawacki-Richter et al. [9] mengelompokkan aplikasi kecerdasan buatan di pendidikan tinggi ke dalam pemodelan dan prediksi, asesmen, personalisasi, serta tutor cerdas. Literatur ini menunjukkan bahwa asesmen dan personalisasi telah banyak dikaji. Oleh karena itu, istilah adaptif pada PAKSALEH ditempatkan sebagai arah penyesuaian rekomendasi berdasarkan profil hasil, tanpa menyamakan rancangan produk dengan sistem kecerdasan buatan atau tes adaptif terkomputerisasi yang telah tervalidasi.

Kesenjangan yang menjadi fokus penelitian ini bersifat kontekstual dan evaluatif: diperlukan rancangan aplikasi K3 maritim yang menghubungkan asesmen dengan tindak lanjut pengembangan serta bukti penilaian kualitas dari pengguna. Kajian Karimi et al. [3] telah membahas adaptasi untuk keselamatan maritim, sehingga penelitian ini tidak mengajukan klaim sebagai sistem adaptif maritim pertama. Ruiz Nepomuceno et al. [10] menunjukkan bahwa arsitektur sistem pembelajaran adaptif perlu dinilai dari sisi kualitas perangkat lunak. Ahmad Faudzi et al. [11] menekankan peran rancangan antarmuka dan evaluasi *usability* pada aplikasi pembelajaran. Dengan demikian, kajian PAKSALEH mengisi kebutuhan dokumentasi rancangan produk dan evaluasi persepsi kualitas pada konteks pengguna yang diteliti, sementara efektivitas pembelajaran dan ketepatan rekomendasi memerlukan penelitian tersendiri.

Pendekatan yang digunakan adalah pengembangan PAKSALEH dengan kerangka ADDIE dan evaluasi menggunakan instrumen yang mengacu pada enam karakteristik ISO 9126. Alur rancangan menghubungkan asesmen, penghitungan skor, pemetaan kompetensi, identifikasi kesenjangan, rekomendasi, asesmen ulang, dan laporan. Kontribusi produk berada pada integrasi alur tersebut untuk kebutuhan K3 maritim. Evaluasi kualitas diperlukan karena kebermanfaatan rancangan bergantung pada kesesuaian fungsi dan pengalaman pengguna. Ali et al. [12] menjelaskan posisi model kualitas perangkat lunak dan hubungan ISO 9126 dengan perkembangan keluarga SQuaRE. ISO 9126 digunakan untuk mempertahankan kesesuaian dengan instrumen penelitian yang tersedia, bukan sebagai pernyataan sertifikasi ataupun sebagai klaim penggunaan standar kualitas paling mutakhir.

Penelitian ini memiliki dua rumusan masalah: (1) bagaimana rancangan dan proses pengembangan aplikasi PAKSALEH sebagai sistem asesmen dan pengembangan kompetensi adaptif pada K3 maritim; dan (2) bagaimana kualitas aplikasi PAKSALEH berdasarkan penilaian pengguna pada aspek *functionality*, *reliability*, *usability*, *efficiency*, *maintainability*, dan *portability*. Tujuannya adalah menyajikan rancangan pengembangan produk dan mendeskripsikan kualitas yang dipersepsikan pengguna. Kontribusi empiris penelitian berupa ringkasan 104 respons evaluasi, distribusi penilaian, dan identifikasi karakteristik yang perlu diprioritaskan dalam penyempurnaan. Batas penelitian ditetapkan pada kualitas menurut pengguna agar hasil evaluasi perangkat lunak tidak ditafsirkan sebagai bukti peningkatan kompetensi atau keselamatan kerja.

2. METODOLOGI PENELITIAN

Penelitian menggunakan pendekatan penelitian dan pengembangan (R&D) dengan kerangka ADDIE. Kerangka ini digunakan untuk mengorganisasi kebutuhan pembelajaran dan rancangan aplikasi; penggunaan ADDIE dalam pengembangan aplikasi pembelajaran vokasional juga dijelaskan oleh Stapa dan Mohammad [13].

Pada PAKSALEH, tahap *analysis* menetapkan kebutuhan asesmen dan tindak lanjut K3 maritim; *design* menyusun alur asesmen, skoring, pemetaan, dan rekomendasi; *development* mengarahkan pembangunan komponen aplikasi; *implementation* menyediakan konteks penggunaan untuk evaluasi; dan *evaluation* merangkum penilaian pengguna serta prioritas penyempurnaan. Uraian proses mengacu pada dokumen rancangan pengembangan. Penilaian empiris dalam artikel ini berasal dari data kuesioner, sehingga uraian tahapan tersebut tidak diperlakukan sebagai hasil audit implementasi setiap fitur.

Data evaluasi terdiri atas 104 baris respons dan 18 butir pernyataan dengan skor 1–5. Tiga butir berurutan mewakili masing-masing karakteristik: F1–F3 (*functionality*), R1–R3 (*reliability*), U1–U3 (*usability*), E1–E3 (*efficiency*), M1–M3 (*maintainability*), dan P1–P3 (*portability*). Pemeriksaan kelengkapan dan rentang skor dilakukan sebelum analisis. Seluruh respons dipertahankan karena tidak ada nilai hilang atau skor di luar rentang. Data semester mencatat 103 responden pada semester 2 dan satu responden pada semester 8; tidak dilakukan pembatasan analisis berdasarkan semester. Nama responden tidak disajikan. Instrumen digunakan sebagai indikator persepsi kualitas, bukan sebagai tes kompetensi. Tidak tersedia bukti validasi konstruk atau validasi ahli dalam bahan penelitian, sehingga interpretasi memperhatikan prinsip kejelasan domain dan kesesuaian butir dengan pengetahuan responden [14].

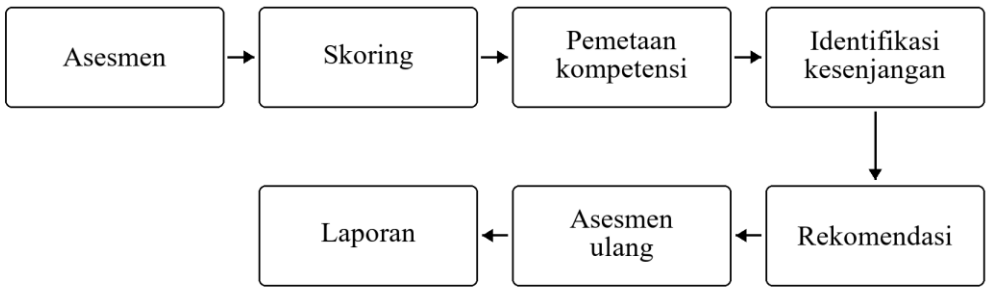
Analisis dilakukan secara deskriptif. Rata-rata karakteristik dihitung dari 312 jawaban (104 responden × 3 butir), sedangkan rata-rata keseluruhan dihitung dari 1.872 jawaban. Setiap karakteristik memiliki bobot sama. Simpangan baku (SB) dihitung dari 104 skor komposit responden, yaitu rata-rata tiga butir per karakteristik atau rata-rata 18 butir untuk keseluruhan, dengan penyebut $n-1$. Persentase frekuensi skor dihitung terhadap jumlah jawaban, bukan jumlah responden. Untuk konsistensi dengan kategori pada rancangan, rata-rata dibulatkan dua desimal sebelum dikategorikan: 1,00–1,80 sangat kurang; 1,81–2,60 kurang; 2,61–3,40 cukup; 3,41–4,20 baik; dan 4,21–5,00 sangat baik. Kategori tersebut merupakan interval deskriptif penelitian, bukan ambang baku ISO. Tidak dilakukan pengujian sebab-akibat atau pengujian perbedaan antarkarakteristik.

3. HASIL DAN PEMBAHASAN

3.1 Rancangan Pengembangan Produk

Rancangan pengembangan PAKSALEH menempatkan aplikasi sebagai penghubung antara pengukuran capaian dan tindak lanjut pengembangan kompetensi. Masukan utama pada rancangan adalah respons asesmen yang dihubungkan dengan indikator kompetensi K3 maritim. Pengolahan hasil diarahkan pada penyajian skor dan pemetaan indikator agar area yang memerlukan penguatan dapat dikenali. Keluaran yang direncanakan berupa rekomendasi pengembangan dan laporan perkembangan setelah asesmen ulang. Hubungan komponen tersebut diringkas pada Gambar 1.

Alur tersebut memperlihatkan bahwa tahap rekomendasi bergantung pada hasil pemetaan kompetensi, sementara asesmen ulang menyediakan masukan untuk pemantauan berikutnya. Unsur adaptif dipahami sebagai penyesuaian tindak lanjut menurut profil capaian pengguna. Dokumen rancangan belum memberikan aturan ambang kompetensi, pemetaan butir asesmen K3, atau ukuran akurasi rekomendasi. Karena itu, hasil pengembangan yang dapat dilaporkan dari bahan penelitian adalah struktur alur dan tujuan integrasinya. Bukti evaluasi pengguna pada bagian berikut melengkapi uraian produk, tetapi tidak memverifikasi ketepatan semua proses internal tersebut.



Gambar 1. Alur konseptual PAKSALEH berdasarkan dokumen rancangan

3.2 Penilaian Kualitas Aplikasi

Dataset evaluasi memuat 104 responden dan 1.872 jawaban lengkap. Skor yang ditemukan adalah 2, 3, 4, dan 5; tidak terdapat skor 1. Rata-rata keseluruhan sebesar 4,35096, atau 4,35 setelah pembulatan, dengan SB skor komposit 0,76. Ringkasan enam karakteristik disajikan pada Tabel 1. Masing-masing karakteristik dihitung dari jumlah butir dan jumlah responden yang sama.

Tabel 1. Penilaian enam karakteristik kualitas aplikasi

Karakteristik	Rata-rata	SB	Kategori
Functionality	4,41	0,78	Sangat baik
Reliability	4,20	0,86	Baik
Usability	4,34	0,79	Sangat baik
Efficiency	4,36	0,79	Sangat baik
Maintainability	4,38	0,79	Sangat baik
Portability	4,42	0,76	Sangat baik
Keseluruhan	4,35	0,76	Sangat baik

SB dihitung dari 104 skor komposit responden. Kategori diterapkan setelah pembulatan dua desimal. Sumber: data evaluasi diolah.

Portability memiliki rata-rata tertinggi (4,42), diikuti *functionality* (4,41), *maintainability* (4,38), *efficiency* (4,36), dan *usability* (4,34). *Reliability* memiliki rata-rata terendah (4,20). Selisih rata-rata *portability* dan *reliability* sebelum pembulatan sebesar 0,21474. Selisih ini dilaporkan secara deskriptif dan tidak diuji sebagai perbedaan yang signifikan. SB antarkarakteristik berkisar 0,76–0,86, dengan nilai terbesar pada *reliability*. Kategori *reliability* mengikuti nilai yang dibulatkan menjadi 4,20; nilai sebelum pembulatan adalah 4,20192.

Rincian penilaian setiap butir ditampilkan pada Tabel 2. Rata-rata butir tertinggi terdapat pada kesesuaian informasi dengan kebutuhan pengguna (F2; 4,45), sedangkan terendah pada stabilitas dalam berbagai kondisi (R1; 4,19). Ketiga butir *reliability* berada pada rentang rata-rata 4,19–4,21. Butir teknis M2 tentang struktur kode atau dokumentasi memperoleh rata-rata 4,40, dan E3 tentang penggunaan memori memperoleh rata-rata 4,38; angka tersebut merupakan respons kuesioner terhadap pernyataan, tanpa pengukuran kode atau memori perangkat.

Tabel 2. Rata-rata setiap butir evaluasi

Kode	Ringkasan pernyataan	Rata-rata
F1	Fitur utama berjalan sesuai harapan	4,41
F2	Informasi sesuai kebutuhan pengguna	4,45
F3	Tidak ada kesalahan proses fitur utama	4,36
R1	Stabil pada berbagai kondisi	4,19
R2	Tidak sering mengalami *crash*	4,20
R3	Menangani kesalahan tanpa kegagalan	4,21
U1	Antarmuka mudah dipahami	4,37
U2	Navigasi mudah dioperasikan	4,32
U3	Panduan penggunaan jelas	4,35
E1	Waktu respons cepat	4,36
E2	Berjalan lancar tanpa hambatan	4,36
E3	Penggunaan memori tidak berlebihan	4,38
M1	Mudah diperbarui atau dimodifikasi	4,36
M2	Struktur kode atau dokumentasi jelas	4,40
M3	*Bug* mudah diidentifikasi dan diperbaiki	4,37
P1	Berjalan pada berbagai perangkat web	4,42
P2	Instalasi mudah	4,41
P3	Tidak ada masalah kompatibilitas antarversi	4,41

Setiap butir memiliki 104 jawaban. Redaksi pernyataan diringkas untuk penyajian. Sumber: data evaluasi diolah.

Distribusi pada Tabel 3 menunjukkan 1.041 jawaban berskor 5 (55,61%), 479 jawaban berskor 4 (25,59%), 320 jawaban berskor 3 (17,09%), dan 32 jawaban berskor 2 (1,71%). Gabungan skor 4–5 berjumlah 1.520 jawaban atau 81,20%. Persentase ini menunjukkan proporsi jawaban butir, sehingga tidak dapat ditafsirkan sebagai 81,20% responden yang seluruh penilaiannya positif. Ringkasan frekuensi memberi informasi tambahan tentang komposisi penilaian yang tidak terlihat dari rata-rata saja.

Tabel 3. Distribusi seluruh jawaban evaluasi

Skor	Frekuensi	Persentase
1	0	0,00%
2	32	1,71%
3	320	17,09%
4	479	25,59%
5	1041	55,61%
Total	1.872	100,00%

Sumber: data evaluasi diolah. Unit frekuensi adalah jawaban butir.

3.3 Pembahasan

Rancangan PAKSALEH menjawab rumusan masalah pertama pada tingkat struktur proses: hasil asesmen diarahkan menjadi pemetaan kompetensi dan dasar tindak lanjut. Posisi ini sejalan dengan asesmen digital yang mendukung umpan balik [4] dan asesmen formatif yang memperhatikan kebutuhan individual [5]. Kontribusinya terletak pada pengorganisasian alur untuk K3 maritim, bukan pengenalan konsep personalisasi yang sepenuhnya baru. Literatur keselamatan maritim telah membahas pemodelan peserta dan adaptasi instruksional [3]. Untuk memperkuat kontribusi pengembangan, dokumentasi berikutnya perlu menghubungkan setiap indikator K3 dengan butir asesmen, aturan diagnosis, dan sumber materi rekomendasi. Hubungan yang dapat ditelusuri ini diperlukan agar rekomendasi memiliki alasan yang jelas bagi pengguna dan pengelola pelatihan.

Rumusan masalah kedua memperoleh jawaban melalui kecenderungan penilaian yang positif pada keenam karakteristik. Rata-rata keseluruhan 4,35 dan dominasi jawaban 4–5 menggambarkan kualitas yang dipersepsikan baik oleh pengguna dalam *dataset*. Hasil ini dapat menjadi dasar untuk melanjutkan penyempurnaan dan uji penggunaan yang lebih terarah. Namun, penilaian persepsi tidak sama dengan pengujian akurasi skoring, keamanan penyimpanan, atau ketepatan diagnosis. Model kualitas memberikan kerangka karakteristik yang harus diterjemahkan menjadi ukuran dan bukti sesuai konteks evaluasi [12]. Oleh karena itu, artikel ini menggunakan istilah kualitas menurut pengguna dan tidak menyimpulkan bahwa PAKSALEH telah memenuhi seluruh persyaratan teknis ISO.

Functionality yang mencapai 4,41 menunjukkan penilaian positif terhadap fungsi utama dan informasi yang ditampilkan. Nilai F2 sebesar 4,45 mendukung interpretasi bahwa informasi aplikasi dipersepsikan sesuai kebutuhan. Untuk sistem yang menghubungkan asesmen dengan rekomendasi, kesesuaian informasi merupakan prasyarat agar hasil dapat dipahami dan ditindaklanjuti. Kajian asesmen adaptif menjelaskan penggunaan informasi kemampuan peserta untuk penyesuaian layanan belajar [7]. Akan tetapi, butir *functionality* pada penelitian ini bersifat umum dan belum menanyakan secara khusus ketepatan setiap rekomendasi. Karena itu, penilaian tinggi dapat mendukung pengembangan fungsi, tetapi belum menjadi bukti validitas pedagogis mekanisme adaptasi.

Usability (4,34) dan *efficiency* (4,36) memberikan gambaran positif mengenai kemudahan interaksi dan respons yang dirasakan pengguna. Kajian Ahmad Faudzi et al. [11] menunjukkan bahwa navigasi, konsistensi, dan bantuan penggunaan menjadi persoalan penting dalam rancangan antarmuka aplikasi pembelajaran. Pada PAKSALEH, navigasi memiliki rata-rata 4,32, lebih rendah daripada kemudahan memahami antarmuka (4,37). Perbedaan kecil ini dapat digunakan untuk memilih fokus observasi berikutnya, misalnya menemukan hasil pemetaan dan membuka rekomendasi. Pengamatan berbasis tugas perlu mencatat keberhasilan, waktu, dan kesalahan pengguna. Ukuran tersebut melengkapi persepsi *efficiency* tanpa menganggap jawaban tentang penggunaan memori sebagai pengukuran sumber daya komputasi.

Portability (4,42) merupakan kekuatan relatif dalam penilaian pengguna. Dalam rancangan layanan digital, akses melalui perangkat yang beragam dapat memperluas kesempatan mengikuti asesmen dan tindak lanjut. Arsitektur sistem pembelajaran adaptif perlu mempertimbangkan kualitas serta kebutuhan pengguna pada lingkungan aksesnya [10]. Namun, *dataset* PAKSALEH tidak memuat jenis perangkat, peramban, atau versi sistem operasi. Dengan demikian, nilai *portability* menunjukkan persepsi terhadap kemudahan penggunaan lintas lingkungan, belum hasil uji kompatibilitas yang terkontrol. Verifikasi berikutnya dapat menggunakan matriks perangkat dan peramban serta skenario pengisian, penyimpanan, dan pembukaan laporan.

Reliability menjadi prioritas penyempurnaan karena rata-ratanya terendah dan variasi skor kompositnya terbesar. Butir stabilitas dalam berbagai kondisi (4,19) menjadi petunjuk awal untuk menentukan skenario pengujian. Pengujian lanjutan dapat mencakup koneksi terputus saat pengisian, permintaan berulang, pemulihan sesi, dan penyimpanan jawaban. Skenario tersebut merupakan rekomendasi pengembangan, bukan kejadian kegagalan yang telah dibuktikan oleh *dataset*. Selain itu, perbedaan kategori baik dan sangat baik pada batas 4,20 perlu dipahami sebagai konsekuensi konvensi pembulatan. Tidak ada dasar untuk menyatakan bahwa *reliability* gagal memenuhi ambang ISO atau berbeda signifikan dari karakteristik lain.

Maintainability (4,38) membutuhkan interpretasi paling hati-hati. Pernyataan tentang struktur kode, dokumentasi, dan kemudahan memperbaiki *bug* memerlukan pengalaman teknis yang tidak dapat diasumsikan dimiliki seluruh responden. Prinsip pengembangan skala menuntut kesesuaian domain dan keterpahaman butir bagi populasi penilai [14]. Nilai tersebut sebaiknya diperlakukan sebagai persepsi atas pernyataan instrumen. Pemeriksaan kode, dokumentasi, pengujian regresi, dan penilaian pengembang akan memberikan bukti yang lebih tepat untuk kualitas internal. Pemisahan instrumen pengguna akhir dan instrumen ahli juga dapat meningkatkan ketepatan pengukuran dalam penelitian berikutnya.

Implikasi pedagogis rancangan terletak pada pemanfaatan informasi hasil untuk tindakan belajar berikutnya. Tinjauan Bayly-Castaneda et al. [6] dan Gligorea et al. [8] mendukung perhatian pada personalisasi, tetapi keduanya tidak membuktikan bahwa seluruh bentuk adaptasi bekerja sama efektif. Kajian Zawacki-Richter et al. [9] juga menunjukkan perlunya hubungan yang lebih kuat dengan perspektif pedagogis. PAKSALEH perlu memperlihatkan kepada pengguna indikator yang belum dikuasai, alasan rekomendasi, dan cara memantau perkembangan. Ngulube dan Ncube [15] menyoroti pentingnya integrasi *learning analytics* yang memperhatikan konteks dan pengalaman pengguna. Dengan demikian, pengembangan laporan dapat diarahkan pada informasi yang mudah ditindaklanjuti, bukan hanya penambahan grafik atau skor.

Penelitian memiliki keterbatasan pada cakupan bukti. Dokumentasi pengembangan belum dilengkapi kode, rekaman pengujian fitur, atau demonstrasi rekomendasi; data evaluasi merupakan laporan diri pada satu *dataset*. Tidak tersedia ukuran kompetensi awal dan akhir, kelompok pembanding, maupun rekaman proses asesmen. Bulut dan Yildirim-Erbasli [16] menjelaskan bahwa analitik asesmen dapat menggunakan data proses untuk memperkuat pemahaman tentang pengukuran digital. Perekaman waktu respons, perubahan jawaban, dan tindakan pengguna dapat menjadi agenda berikutnya apabila dikumpulkan dengan tata kelola yang sesuai. Studi lanjutan perlu menggabungkan validasi ahli K3, pengujian perangkat lunak, evaluasi instrumen, dan desain pengukuran kompetensi. Dengan batas tersebut, hasil penelitian saat ini memberikan bukti awal persepsi kualitas dan arah penyempurnaan, bukan kesimpulan tentang peningkatan keselamatan kerja.

4. KESIMPULAN

Pengembangan aplikasi PAKSALEH melalui kerangka ADDIE menghasilkan rancangan yang mengintegrasikan asesmen K3 maritim, penghitungan skor, pemetaan kompetensi, identifikasi kesenjangan, rekomendasi pengembangan, asesmen ulang, dan pelaporan. Integrasi tersebut menjadi dasar penyesuaian tindak lanjut pengembangan kompetensi berdasarkan profil capaian individu. Evaluasi terhadap 104 responden menggunakan 18 butir pernyataan menunjukkan bahwa kualitas aplikasi menurut pengguna berada pada kategori sangat baik, dengan rata-rata keseluruhan 4,35 dan proporsi jawaban berskor 4–5 sebesar 81,20%. Nilai rata-rata functionality sebesar 4,41, reliability 4,20, usability 4,34, efficiency 4,36, maintainability 4,38, dan portability 4,42. Lima karakteristik termasuk kategori sangat baik, sedangkan reliability termasuk kategori baik. Portability menjadi kekuatan relatif aplikasi dalam persepsi pengguna, sementara reliability menjadi prioritas penyempurnaan, terutama pada stabilitas sistem. Temuan ini memberikan dasar bagi penyempurnaan rancangan dan kualitas aplikasi, dengan batasan bahwa penilaian pengguna belum membuktikan ketepatan rekomendasi adaptif, kualitas teknis internal, maupun peningkatan kompetensi dan keselamatan kerja. Penelitian selanjutnya disarankan mencakup validasi indikator kompetensi dan aturan rekomendasi oleh ahli K3 maritim, pengujian teknis serta validasi instrumen, dan pengukuran perubahan kompetensi sebelum–sesudah penggunaan dengan kelompok pembanding.

DAFTAR PUSTAKA

[1] M. H. Dewan, R. Godina, M. R. K. Chowdhury, C. W. M. Noor, W. M. N. Wan Nik, and M. Man, “Immersive and Non-Immersive Simulators for the Education and Training in Maritime Domain—A Review,” *J. Mar. Sci. Eng.*, vol. 11, no. 1, 2023, doi: 10.3390/jmse11010147.

[2] R. P. de Oliveira, G. Carim Junior, B. Pereira, D. Hunter, J. Drummond, and M. Andre, “Systematic Literature Review on the Fidelity of Maritime Simulator Training,” *Educ. Sci.*, vol. 12, no. 11, 2022, doi: 10.3390/educsci12110817.

[3] E. Karimi, J. Smith, R. Billard, and B. Veitch, “AI-based adaptive instructional systems for maritime safety training: a systematic literature review,” *Discov. Artif. Intell.*, vol. 4, no. 1, 2024, doi: 10.1007/s44163-024-00153-0.

[4] A. Jurāne-Brēmane, “Digital Assessment in Technology-Enriched Education: Thematic Review,” *Educ. Sci.*, vol. 13, no. 5, 2023, doi: 10.3390/educsci13050522.

[5] B. P. Solis Trujillo, D. Velarde-Camaqui, C. A. Gonzales Nuñez, E. V. Castillo Silva, and M. del P. Gonzalez Said de la Oliva, “The current landscape of formative assessment and feedback in graduate studies: a systematic literature review,” *Front. Educ.*, vol. 10, 2025, doi: 10.3389/feduc.2025.1509983.

[6] K. Bayly-Castaneda, M. S. Ramirez-Montoya, and A. Morita-Alexander, “Crafting personalized learning

paths with AI for lifelong learning: a systematic literature review," *Front. Educ.*, vol. 9, no. August, pp. 1–12, 2024, doi: 10.3389/feduc.2024.1424386.

[7] A. Ihichr, O. Oustous, Y. El Bouzekri, and A. A. Lahcen, "A Systematic Review on Assessment in Adaptive Learning: Theories, Algorithms and Techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 7, pp. 855–868, 2024, doi: 10.14569/IJACSA.2024.0150785.

[8] I. Gligorea, M. Cioca, R. Oancea, A. T. Gorski, H. Gorski, and P. Tudorache, "Adaptive Learning Using Artificial Intelligence in e-Learning: A Literature Review," *Educ. Sci.*, vol. 13, no. 12, 2023, doi: 10.3390/educsci13121216.

[9] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education – where are the educators?," *Int. J. Educ. Technol. High. Educ.*, vol. 16, no. 1, 2019, doi: 10.1186/s41239-019-0171-0.

[10] A. Ruiz Nepomuceno, E. López Domínguez, S. Domínguez Isidro, M. A. Medina Nieto, A. Meneses-Viveros, and J. de la Calleja, "Software Architectures for Adaptive Mobile Learning Systems: A Systematic Literature Review," *Appl. Sci.*, vol. 14, no. 11, 2024, doi: 10.3390/app14114540.

[11] M. Ahmad Faudzi, Z. Che Cob, R. Omar, S. A. Sharudin, and M. Ghazali, "Investigating the User Interface Design Frameworks of Current Mobile Learning Applications: A Systematic Review," *Educ. Sci.*, vol. 13, no. 1, 2023, doi: 10.3390/educsci13010094.

[12] M. A. Ali, N. K. Yap, A. A. A. Ghani, H. Zulzalil, N. I. Admodisastro, and A. A. Najafabadi, "A Systematic Mapping of Quality Models for AI Systems, Software and Components," *Appl. Sci.*, vol. 12, no. 17, 2022, doi: 10.3390/app12178700.

[13] M. A. Stapa and N. Mohammad, "The Use of Addie Model for Designing Blended Learning Application at Vocational Colleges in Malaysia," *Asia-Pacific J. Inf. Technol. Multimed.*, vol. 08, no. 01, pp. 49–62, 2019, doi: 10.17576/apjitm-2019-0801-05.

[14] G. O. Boateng, T. B. Neilands, E. A. Frongillo, H. R. Melgar-Quiñonez, and S. L. Young, "Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer," *Front. Public Heal.*, vol. 6, no. June, pp. 1–18, 2018, doi: 10.3389/fpubh.2018.00149.

[15] P. Ngulube and M. M. Ncube, "Leveraging Learning Analytics to Improve the User Experience of Learning Management Systems in Higher Education Institutions," *Inf.*, vol. 16, no. 5, pp. 1–20, 2025, doi: 10.3390/info16050419.

[16] O. Bulut and S. N. Yildirim-Erbasli, "Assessment Analytics in Digital Assessments," *Encyclopedia*, vol. 6, no. 4, pp. 1–18, 2026, doi: 10.3390/encyclopedia6040081.