

Pemetaan Metode Evaluasi, Objek Evaluasi, dan Aspek UI/UX dalam Studi Komparatif: Sebuah Systematic Literature Review

Mapping UI/UX Evaluation Methods, Evaluation Objects, and Measured Aspects in Comparative Studies: A Systematic Literature Review

Bachtiar Mujaddidi
Fakultas Ilmu Komputer
Universitas Amikom Purwokerto
Purwokerto, Indonesia
bahctiarmujaddidi@gmail.com

Berlilana
Fakultas Ilmu Komputer
Universitas Amikom Purwokerto
Purwokerto, Indonesia
berli@amikompurwokerto.ac.id

Purwadi
Fakultas Ilmu Komputer
Universitas Amikom Purwokerto
Purwokerto, Indonesia
purwadi@amikompurwokerto.ac.id

Diterima : Juni 2026
Disetujui : Juni 2026
Dipublikasi : Juli 2026

Abstrak— Evaluasi UI/UX merupakan aspek penting dalam memastikan kualitas interaksi antara pengguna dan sistem, terutama dalam konteks meningkatnya kompleksitas aplikasi digital. Namun, beragamnya metode evaluasi yang tersedia seringkali menyebabkan perbedaan hasil dan interpretasi dalam menilai performa sistem dan kepuasan pengguna. Penelitian ini bertujuan untuk mengkaji dan membandingkan pendekatan evaluasi UI/UX yang digunakan dalam studi komparatif. Berbeda dengan penelitian tinjauan sebelumnya yang umumnya berfokus pada domain aplikasi tertentu, persepsi pengguna, atau tren teknologi, penelitian ini secara khusus memetakan hubungan antara metode evaluasi UI/UX, objek evaluasi yang dibandingkan, serta dimensi pengalaman pengguna yang diukur dalam studi komparatif. Metode yang digunakan adalah *Systematic Literature Review* (SLR) dengan mengikuti panduan PRISMA. Pencarian literatur dilakukan pada database Scopus, dan melalui proses seleksi diperoleh 13 artikel yang memenuhi kriteria inklusi. Hasil analisis menunjukkan bahwa pendekatan evaluasi UI/UX dalam studi komparatif didominasi oleh perbandingan teknik atau pendekatan interaksi sebanyak 61,5% (8 artikel), diikuti oleh perbandingan performa software sebesar 23,1% (3 artikel), serta perbandingan metode evaluasi sebesar 15,4% (2 artikel). Evaluasi pada studi-studi tersebut umumnya menggunakan kombinasi metrik performa, persepsi pengguna, serta indikator pengalaman lainnya seperti beban kognitif dan respons fisiologis. Temuan ini menunjukkan bahwa evaluasi UI/UX dalam studi komparatif tidak hanya digunakan untuk membandingkan metode evaluasi, tetapi juga untuk menilai berbagai alternatif desain interaksi dan sistem berdasarkan beragam dimensi pengalaman pengguna. Kontribusi utama penelitian ini adalah menyediakan sintesis sistematis mengenai hubungan antara metode evaluasi, objek evaluasi, dan indikator UI/UX yang digunakan dalam studi komparatif. Hasil penelitian ini memberikan pemetaan yang lebih komprehensif dibandingkan tinjauan sebelumnya dan dapat menjadi referensi bagi peneliti maupun praktisi dalam merancang evaluasi UI/UX yang lebih tepat dan konsisten.

Kata Kunci— antarmuka pengguna; pengalaman pengguna; Human-Computer Interaction; Usability Testing; User-Centered Design; Evaluation Framework.

Abstract— UI/UX evaluation is a crucial aspect in ensuring the quality of interaction between users and systems, particularly in the context of increasingly complex digital applications. However, the wide variety of available evaluation methods often leads to differences in results and interpretations when assessing system performance and user satisfaction. This study aims to review and compare UI/UX evaluation approaches used in comparative studies. Unlike previous review studies that mainly focused on specific application domains, user perceptions, or technological trends, this study specifically maps the relationship between UI/UX evaluation methods, evaluated objects, and user experience dimensions measured in comparative studies. The method employed is a systematic literature review (SLR) following the PRISMA guidelines. Literature was retrieved from the Scopus database, and through the selection process, 13 articles met the inclusion criteria. The analysis revealed that UI/UX evaluation approaches in comparative studies are predominantly focused on comparing interaction techniques or interaction approaches, accounting for 61.5% (8 articles) of the selected studies. This is followed by comparisons of software performance, which represent 23.1% (3 articles), and comparisons of evaluation methods, which account for 15.4% (2 articles). Evaluations generally employ a combination of performance metrics, user perception measures, and additional experiential indicators such as cognitive workload and physiological responses. The findings show that each method or system demonstrates strengths in specific evaluation dimensions, highlighting the need for a multidimensional and multi-method evaluation approach to obtain a more comprehensive understanding of UI/UX. The main contribution of this study is the development of a systematic synthesis that links evaluation methods, evaluation objects, and UI/UX indicators employed in comparative studies. The findings provide a broader understanding of current evaluation practices and may serve as a reference for researchers and practitioners in designing more appropriate and consistent UI/UX evaluation strategies.

Keywords— *user interface; user experience; Human–Computer Interaction; Usability Testing; User-Centered Design; Evaluation Framework.*

I. PENDAHULUAN

Penggunaan perangkat lunak (*software*) telah menjadi bagian yang tidak terpisahkan dari berbagai aktivitas manusia sehari-hari, seperti dalam bidang pendidikan [1], bisnis [2], hingga kesehatan [3]. Seiring dengan meningkatnya jumlah dan kompleksitas aplikasi, pengguna kini memiliki banyak pilihan *software* yang menawarkan fungsi serupa. Kondisi ini menjadikan kualitas interaksi antara pengguna dan sistem sebagai faktor pembeda utama dalam keberhasilan adopsi dan keberlanjutan penggunaan *software*. Dalam konteks tersebut, evaluasi berbasis respon pengguna menjadi aspek penting karena menentukan sejauh mana pengguna dapat berinteraksi dengan sistem secara efektif, efisien, dan memuaskan. *User Interface* (UI) dan *User Experience* (UX) memegang peranan penting dalam menentukan penerimaan suatu *software*, karena keduanya memengaruhi kualitas interaksi antara pengguna dan sistem.

Fokus UI adalah elemen visual dan komponen yang digunakan pengguna untuk berinteraksi dengan sistem, sedangkan UX mencakup keseluruhan pengalaman pengguna, termasuk persepsi, emosi, dan respons pengguna selama dan setelah penggunaan sistem. Desain UI/UX yang baik berperan penting dalam meningkatkan kualitas interaksi antara pengguna dan sistem [4]. Namun demikian, kualitas UI/UX tidak dapat dinilai hanya berdasarkan asumsi pengembang, sehingga evaluasinya harus melalui penerapan metode yang sistematis. Evaluasi UI/UX menjadi tahapan yang krusial dalam pengembangan *software* dengan tujuan mengukur sejauh mana sistem memenuhi kebutuhan pengguna, baik dari sisi fungsional maupun pengalaman penggunaan.

Berbagai metode evaluasi UI/UX telah dikembangkan dan digunakan dalam penelitian maupun praktik. Pendekatan tersebut meliputi evaluasi berbasis kuesioner, metode kualitatif seperti wawancara dan observasi, evaluasi berbasis tugas, serta metode inspeksi seperti heuristic evaluation dan *cognitive walkthrough*. Selain itu, terdapat pula pendekatan eksperimen seperti A/B testing. Di antara berbagai pendekatan tersebut, *usability testing* mengevaluasi aspek yang fundamental dan paling sering digunakan sebagai dasar evaluasi UI/UX karena menekankan kemudahan dan kejelasan interaksi pengguna dengan sistem dalam mencapai tujuan tertentu [5].

Usability testing dapat mengukur tingkat efektivitas, efisiensi, dan kepuasan pengguna yang mencakup berbagai komponen seperti apakah sistem yang digunakan mudah dipelajari, mudah diingat dan minim *error* [6]. Namun konsep tradisional *usability* telah berevolusi menjadi ide yang lebih luas yang juga mencakup pengalaman pengguna seperti evaluasi dalam aspek emosional, kognitif, ataupun respon fisik [7]. Oleh karena itu evaluasi UI/UX yang lebih komprehensif berperan penting sebagai jembatan dalam mengukur sejauh mana sistem memungkinkan pengguna mencapai tujuannya. Buruknya penilaian terhadap UI/UX sebuah sistem mengindikasikan kegagalan aspek fungsional maupun non-fungsional sebuah sistem dan mengakibatkan ketidakpuasan pengguna.

Terdapat banyak *framework* untuk mengevaluasi performa UI/UX yang mengukur berbagai aspek dan atribut aplikasi. Nakamura et al. menunjukkan bahwa perbedaan

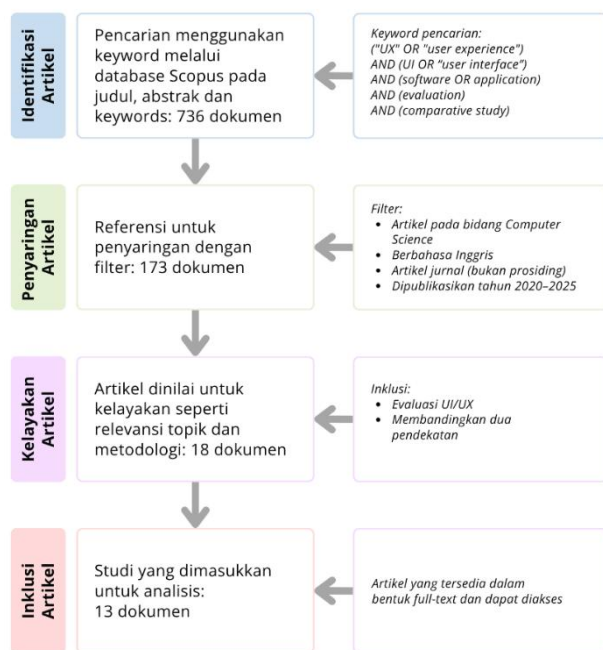
metode evaluasi UI/UX dapat menghasilkan interpretasi yang berbeda meskipun diterapkan pada sistem yang sama [5]. Temuan ini menegaskan bahwa hasil evaluasi dapat dipengaruhi oleh pendekatan evaluasi yang digunakan. Hal ini mengindikasikan perlunya kajian untuk memahami perbedaan tersebut secara objektif terhadap berbagai pendekatan evaluasi lintas studi.

Beberapa studi perbandingan telah dilakukan, seperti studi oleh Theodorou et al. menyajikan studi perbandingan solusi navigasi untuk pengguna tunanetra, dengan *usability* dan UX digunakan sebagai dimensi evaluasi sistem. Namun, fokus studi ini berada pada evaluasi sistem dalam domain aplikasi spesifik, bukan pada perbandingan pendekatan evaluasi secara komparatif [8]. Sandesara et al. melakukan *pilot survey* untuk mengkaji faktor-faktor UI/UX pada aplikasi *mobile* berdasarkan persepsi pengguna melalui kuesioner. Studi ini memberikan gambaran mengenai aspek UI/UX yang dianggap penting oleh pengguna. Namun survei ini tidak membahas pendekatan evaluasi dalam konteks studi komparatif [9]. Sementara itu, Hu et al. menyajikan *review* mengenai penggunaan *virtual reality* dalam pelatihan interpreting, dengan *usability* dan UX diidentifikasi sebagai evaluasi utama. Meskipun demikian, fokus studi ini adalah pemetaan tren dan aplikasi VR, bukan pada perbandingan pendekatan evaluasi UI/UX secara spesifik [10].

Mengingat banyaknya metode dan pendekatan evaluasi UI/UX yang tersedia, diperlukan kajian literatur yang secara sistematis menelaah bagaimana sistem dievaluasi dalam studi komparatif UI/UX. Berbeda dengan penelitian sebelumnya yang umumnya berfokus pada evaluasi aplikasi tertentu, persepsi pengguna, atau tren penerapan teknologi, penelitian ini memusatkan perhatian pada pendekatan evaluasi itu sendiri sebagai objek kajian utama. Keunikan penelitian ini terletak pada upayanya untuk memetakan dan membandingkan berbagai pendekatan evaluasi sistem yang digunakan dalam studi komparatif UI/UX, termasuk dimensi *usability*, *user experience*, metode pengukuran, instrumen yang digunakan, serta indikator keberhasilan evaluasi. Melalui *systematic literature review*, penelitian ini diharapkan memberikan kontribusi teoretis berupa kerangka sintesis yang menghubungkan metode evaluasi dengan aspek UI/UX yang diukur dan konteks penggunaannya. Penelitian ini tidak hanya mengidentifikasi pendekatan yang dominan, tetapi juga menjelaskan pola pemilihan metode, kesesuaian penggunaannya, serta potensi keterbatasannya dalam menghasilkan interpretasi evaluasi UI/UX yang berbeda. Hasil kajian ini diharapkan dapat menjadi dasar konseptual bagi peneliti dan praktisi dalam memilih pendekatan evaluasi UI/UX yang lebih tepat dan konsisten pada penelitian maupun pengembangan selanjutnya.

II. METODE

Studi ini menggunakan metodologi *Systematic Literature Review* (SLR) yang mengikuti panduan *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) [11]. Proses perencanaan, pelaksanaan, dan pelaporan SLR mengacu pada pedoman Kitchenham dan Charters [12] serta *Software Engineering Guidelines for Reporting Secondary Studies* (SEGRESS) [13]. Pendekatan ini digunakan untuk memastikan proses penelusuran, seleksi, dan sintesis literatur dilakukan secara sistematis, transparan, dan dapat direplikasi. Diagram alur metode penelitian ditunjukkan pada Gambar 1.



Gambar 1. Diagram PRISMA seleksi literatur untuk perbandingan evaluasi UI/UX

Penelusuran literatur dilakukan mengikuti tahapan PRISMA yang meliputi identifikasi, penyaringan, penilaian kelayakan, dan inklusi studi. Pada tahap identifikasi, pencarian dilakukan melalui basis data *Scopus* pada judul, abstrak, dan kata kunci. *Search string* yang digunakan adalah sebagai berikut:

TITLE-ABS-KEY(
("user interface" OR UI OR "user experience" OR UX)
AND
(evaluation OR usability OR "user evaluation")
AND
(comparative OR comparison OR benchmark)
)

Pencarian dibatasi pada artikel jurnal berbahasa Inggris yang dipublikasikan pada periode 2020–2025 dan termasuk dalam bidang Computer Science, sehingga diperoleh dokumen awal sebanyak 736. Selanjutnya, pada tahap penyaringan, dokumen diseleksi menggunakan filter yang menghasilkan 173 dokumen. Pada tahap kelayakan, artikel ditinjau lebih lanjut berdasarkan relevansi topik dan metodologi penelitian sehingga diperoleh 18 dokumen. Pada tahap kelayakan, sebanyak 18 artikel ditelaah secara penuh (*full-text review*). Hasil penelaahan menunjukkan bahwa 5 artikel tidak memenuhi kriteria inklusi, yaitu karena tidak berfokus pada studi komparatif UI/UX ($n = 2$), tidak menjelaskan pendekatan evaluasi UI/UX secara memadai ($n = 1$), hanya mengevaluasi aspek teknis sistem tanpa mempertimbangkan UI/UX ($n = 1$), serta tidak menyediakan informasi metodologi yang cukup untuk ekstraksi data ($n = 1$). Sehingga didapatkan 13 artikel memenuhi seluruh kriteria inklusi dan digunakan pada tahap sintesis dan analisis. Setelah tahap seleksi berdasarkan kriteria inklusi dan eksklusi, artikel yang memenuhi syarat dievaluasi lebih lanjut melalui proses *Quality Assessment (QA)*. Penilaian ini

dilakukan untuk memastikan bahwa studi yang digunakan memiliki kualitas metodologis yang memadai, relevan dengan tujuan penelitian, serta menyediakan informasi yang cukup untuk mendukung proses ekstraksi data dan sintesis temuan. Proses QA dilakukan berdasarkan pedoman *systematic literature review* dan *secondary study* yang menekankan pentingnya evaluasi kualitas studi untuk meningkatkan validitas, transparansi, dan reliabilitas hasil sintesis penelitian [12], [13].

Empat kriteria digunakan dalam proses penilaian kualitas sebagaimana ditunjukkan pada Tabel 1. Setiap kriteria diberi skor 1 jika terpenuhi sepenuhnya, 0.5 jika terpenuhi sebagian, dan 0 jika tidak terpenuhi. Total skor maksimum yang dapat diperoleh setiap artikel adalah 4.

TABEL 1. KRITERIA QUALITY ASSESSMENT

Kode	Kriteria Penilaian
QA1	Apakah tujuan penelitian dijelaskan dengan jelas?
QA2	Apakah metode evaluasi UI/UX dijelaskan secara rinci dan dapat dipahami?
QA3	Apakah karakteristik partisipan dan prosedur evaluasi dijelaskan secara memadai?
QA4	Apakah hasil evaluasi dilaporkan secara lengkap dan dapat digunakan untuk proses ekstraksi data?

Artikel dengan skor total minimal 3 dianggap memiliki kualitas yang memadai dan dapat digunakan dalam proses sintesis. Hasil penilaian kualitas terhadap 13 artikel yang memenuhi kriteria inklusi ditunjukkan pada Tabel 2.

TABEL 2. HASIL QUALITY ASSESSMENT ARTIKEL TERPILIH

Publikasi	QA1	QA2	QA3	QA4	Total
[5]	1	1	1	1	4
[7]	1	1	1	1	4
[15]	1	1	1	1	4
[16]	1	1	0.5	1	3.5
[17]	1	1	1	1	4
[18]	1	1	1	1	4
[19]	1	1	1	1	4
[20]	1	1	0.5	1	3.5
[21]	1	1	1	1	4
[22]	1	1	1	1	4
[23]	1	1	1	1	4
[24]	1	1	1	1	4
[25]	1	1	1	1	4

Berdasarkan hasil *Quality Assessment* pada Tabel 2, seluruh artikel memperoleh skor di atas batas minimum yang telah ditetapkan (≥ 3), sehingga dinilai layak untuk digunakan dalam proses ekstraksi data dan sintesis temuan. Sebagian besar artikel memperoleh skor tinggi karena menjelaskan tujuan penelitian, metode evaluasi, karakteristik partisipan,

prosedur penelitian, serta hasil evaluasi secara memadai. Beberapa artikel memperoleh skor 3.5 karena informasi mengenai karakteristik partisipan atau prosedur evaluasi tidak dijelaskan secara rinci, meskipun masih menyediakan informasi yang cukup untuk mendukung proses ekstraksi data. Secara keseluruhan, hasil ini menunjukkan bahwa studi yang digunakan memiliki kualitas metodologis yang baik dan relevan dengan tujuan penelitian.

Setelah artikel terpilih, dilakukan proses ekstraksi data untuk menjawab pertanyaan penelitian/*research question* (RQ) sebagai berikut:

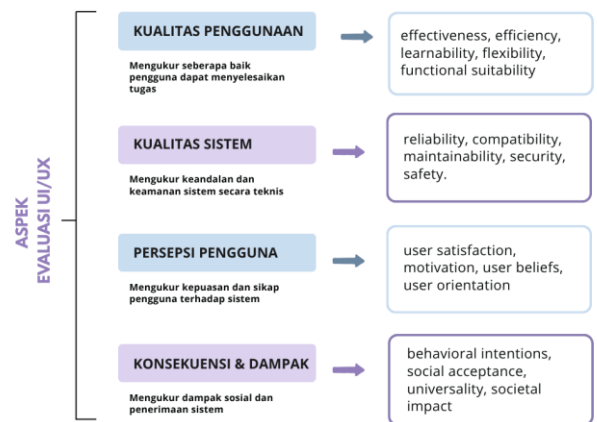
- RQ1: Apa saja objek yang dibandingkan dalam studi komparatif UI/UX dan bagaimana evaluasi UI/UX digunakan dalam perbandingan tersebut? Pertanyaan ini bertujuan mengidentifikasi jenis pendekatan evaluasi yang digunakan dalam penelitian sebelumnya.
- RQ2: Bagaimana prosedur evaluasi UI/UX dilakukan dalam masing-masing studi? Pertanyaan ini mencakup konteks evaluasi, teknik pengumpulan data dan partisipan penelitian.
- RQ3: Aspek UI/UX apa yang dievaluasi? Pertanyaan ini berfokus pada dimensi evaluasi seperti efektivitas, efisiensi, dan kepuasan pengguna dan lain-lain.
- RQ4: Bagaimana hasil perbandingan berdasarkan pendekatan evaluasi UI/UX? Pertanyaan ini bertujuan mengidentifikasi hasil perbandingan berdasarkan pendekatan evaluasi pada RQ1.

Sintesis data dilakukan menggunakan pendekatan *narrative synthesis* [14] dengan mengelompokkan hasil ekstraksi berdasarkan pertanyaan penelitian (RQ). Tahap pertama dilakukan dengan mengidentifikasi objek yang dibandingkan dalam setiap studi komparatif UI/UX. Selanjutnya dilakukan identifikasi prosedur evaluasi yang digunakan, aspek UI/UX yang diukur, serta hasil perbandingan yang dilaporkan pada masing-masing penelitian. Temuan-temuan yang memiliki karakteristik serupa kemudian dikelompokkan ke dalam kategori yang sama untuk mengidentifikasi pola dominan, persamaan, dan perbedaan antar studi. Hasil pengelompokan tersebut digunakan untuk menyusun sintesis yang sistematis mengenai hubungan antara metode evaluasi, objek evaluasi, dan aspek UI/UX yang diukur dalam studi komparatif. Pendekatan ini memungkinkan interpretasi yang lebih terstruktur serta mendukung transparansi dan konsistensi dalam proses sintesis literatur [13], [14].

III. HASIL DAN PEMBAHASAN

Berbagai konsep dan aspek evaluasi UI/UX telah dikemukakan dalam literatur *Human-Computer Interaction* (HCI). *Usability* merupakan salah satu komponen utama dalam evaluasi tersebut, yang menurut International Organization for Standardization (ISO) 2019 didefinisikan sebagai tingkat efektivitas, efisiensi, dan kepuasan pengguna dalam mencapai tujuan tertentu pada konteks penggunaan tertentu. Sementara itu, Nielsen (1992) memandang *usability* sebagai atribut kualitas UI yang mencakup lima komponen utama, yaitu *learnability*, *efficiency*, *memorability*, *errors*, dan *satisfaction*. Dalam konteks evaluasi UI/UX yang lebih luas, aspek yang dinilai tidak hanya terbatas pada *usability*, tetapi juga mencakup persepsi pengguna, pengalaman

emosional, serta respons selama interaksi. Gambar 2 menunjukkan berbagai aspek evaluasi yang digunakan dalam studi UI/UX, yang dapat dikelompokkan ke dalam empat kategori, yaitu kualitas penggunaan, kualitas sistem, persepsi pengguna, dan konsekuensi penggunaan [6].



Gambar 2. Aspek evaluasi dalam studi UI/UX

Meskipun berbagai pendekatan evaluasi telah digunakan dalam beragam aplikasi, studi yang secara sistematis meninjau bagaimana pendekatan tersebut diterapkan dalam studi komparatif masih terbatas. Analisis diperlukan untuk memahami perbedaan cakupan aspek, fokus evaluasi, serta implikasi dari pendekatan evaluasi yang digunakan. Oleh karena itu, bagian ini menyajikan hasil analisis berdasarkan studi-studi komparatif yang meninjau penerapan pendekatan evaluasi UI/UX dalam berbagai literatur.

A. Objek yang Dibandingkan dalam Studi Komparatif UI/UX

Berdasarkan hasil ekstraksi data dari 13 artikel yang dianalisis, ditemukan bahwa studi komparatif UI/UX membandingkan berbagai objek evaluasi yang dapat dikelompokkan ke dalam tiga kategori utama, yaitu perbandingan metode evaluasi, perbandingan teknik atau pendekatan interaksi, serta perbandingan software atau sistem yang dievaluasi. Pengelompokan ini digunakan untuk memahami bagaimana evaluasi UI/UX diterapkan dalam berbagai konteks penelitian komparatif. Tabel 3 menunjukkan distribusi objek yang dibandingkan dalam studi komparatif UI/UX yang dianalisis.

TABEL 3. PENGELOMPOKAN STUDI BERDASARKAN OBJEK YANG DIBANDINGKAN DALAM STUDI KOMPARATIF UI/UX

Objek yang Dibandingkan	Publikasi	Jumlah
Perbandingan metode evaluasi	[5], [15]	2
Perbandingan teknik/interaksi	[7], [16]–[22]	8
Perbandingan performa <i>software</i>	[23]–[25]	3

Berdasarkan Tabel 3, sebagian besar studi komparatif UI/UX berfokus pada perbandingan teknik atau pendekatan interaksi (61,5%), diikuti oleh perbandingan performa software (23,1%), sedangkan studi yang secara langsung

membandingkan metode evaluasi masih relatif terbatas (15,4%).

Kategori pertama mencakup studi yang secara langsung membandingkan metode evaluasi. Dari 13 artikel yang dianalisis, hanya dua studi yang termasuk dalam kategori ini. Nakamura et al. membandingkan metode pengumpulan data dalam evaluasi UX, yaitu AttrakDiff, UEQ, sementara metode *Sentence Completion* digunakan sebagai pelengkap untuk menilai konsistensi hasil evaluasi pada sistem yang sama. Hasil penelitian menunjukkan bahwa metode yang berbeda dapat menghasilkan interpretasi yang berbeda terhadap pengalaman pengguna [5]. Sementara itu, Kuric et al. membandingkan beberapa varian *tree testing* dalam mengevaluasi struktur informasi UI, dan menemukan adanya perbedaan hasil antar metode yang digunakan. Temuan pada kedua studi ini menunjukkan bahwa pemilihan metode evaluasi dapat mempengaruhi hasil evaluasi yang diperoleh [15].

Sebagian besar artikel termasuk dalam kategori kedua, mencakup studi yang membandingkan teknik atau pendekatan interaksi. Studi-studi tersebut membandingkan berbagai pendekatan, seperti perbandingan *unimodal* dan *multimodal interaction* [16], perbandingan *natural interaction* dengan teknik tradisional [17], perbandingan perangkat input *Virtual Reality* (VR) [18], [20], [22], serta perbandingan interaksi berbasis *gesture*, *controller*, maupun *Augmented Reality* (AR) [19], [7], [20]. Selain itu, beberapa penelitian juga membandingkan fitur dalam sistem untuk melihat pengaruhnya terhadap efektivitas, efisiensi, dan kepuasan pengguna [18]. Studi-studi tersebut menunjukkan bahwa evaluasi UI/UX dapat digunakan sebagai indikator utama dalam menilai efektivitas berbagai teknik interaksi yang dibandingkan.

Kategori ketiga mencakup studi yang membandingkan *software* atau sistem yang memiliki fungsi serupa. *Software* yang dibandingkan mencakup perangkat lunak desain UI [21], aplikasi *mind mapping* [20] dan *platform* pembelajaran daring [22]. Hasil pada kategori ini menunjukkan bahwa evaluasi UI/UX digunakan untuk mengidentifikasi kelebihan dan keterbatasan masing-masing sistem berdasarkan berbagai dimensi pengalaman pengguna.

Secara keseluruhan, hasil analisis menunjukkan bahwa studi komparatif UI/UX lebih banyak menggunakan evaluasi sebagai alat untuk membandingkan teknik interaksi maupun *software* dibandingkan untuk membandingkan metode evaluasi itu sendiri. Temuan ini menunjukkan bahwa evaluasi UI/UX lebih sering dimanfaatkan sebagai metrik penilaian dalam membandingkan alternatif desain dan sistem daripada sebagai objek kajian utama.

Masih terbatasnya jumlah studi yang secara langsung membandingkan metode evaluasi menunjukkan adanya peluang penelitian lanjutan untuk mengkaji karakteristik, sensitivitas, kelebihan, dan keterbatasan berbagai metode evaluasi UI/UX dalam konteks yang berbeda. Kajian semacam ini penting untuk mendukung pemilihan metode evaluasi yang lebih tepat serta meningkatkan konsistensi interpretasi hasil evaluasi dalam penelitian maupun praktik pengembangan sistem.

B. Prosedur Evaluasi

Prosedur evaluasi UI/UX pada studi yang dianalisis menunjukkan variasi dalam konteks evaluasi, teknik

pengumpulan data, karakteristik partisipan. Perbedaan tersebut mencerminkan beragam pendekatan dalam menerapkan evaluasi pada studi komparatif UI/UX. Ringkasan prosedur evaluasi pada masing-masing studi disajikan pada Tabel 4.

TABEL 4. TEKNIK PENGUMPULAN DATA DAN KARAKTERISTIK EVALUASI UI/UX

Publ ikasi	Tah un	Konteks Evaluasi	Teknik Pengumpulan Data	Partisipa n
[5]	2021	Evaluasi longitudinal UX aplikasi <i>online judging</i>	Kuesioner UX (AttrakDiff, UEQ, Sentence Completion)	68 mahasiswa
[16]	2021	Perbandingan dua desain interaksi pada aplikasi <i>remote control</i> bagi lansia	<i>Task completion</i> (time, error rate, satisfaction), kuesioner (SUS)	10 lansia
[22]	2021	Perbandingan teknik <i>locomotion</i> VR: <i>joystick</i> , <i>teleport</i> dan <i>leaning</i>	<i>Task performance</i> (time, accuracy), kuesioner (SSQ, IPQ, SUS, DAQ)	75 pengguna
[7]	2022	Perbandingan modalitas interaksi terhadap <i>usability</i> dan <i>sense of presence</i> pada <i>Gesture-based</i> dan <i>controller bundled</i>	Kuesioner (SUS, UMUX, PQ)	84 pengguna
[19]	2022	Perbandingan penggunaan sistem UI natural (pelacak mata dan isyarat tangan) dengan UI tradisional (<i>keyboard</i>) pada sistem VR simulasi penerbangan drone.	<i>Task performance</i> (time, accuracy), kuesioner	59 pengguna
[17]	2022	Perbandingan teknik interaksi tradisional (<i>mouse & GUI</i>) dengan <i>natural finger interaction</i> (VR)	<i>Task performance</i> (time, accuracy), kuesioner (SUS)	10 pengguna
[24]	2022	Perbandingan <i>Software</i> desain UI: Sketch, Adobe XD dan Figma	<i>Task performance</i> (time, accuracy), kuesioner (PSSUQ, ASQ)	17 mahasiswa
[18]	2022	Perbandingan empat perangkat input VR (<i>Controllers</i> , <i>VR Ink</i> , <i>Data Gloves</i> , <i>Craniotome</i>)	<i>Task performance</i> (time, precision), kuesioner (NASA-TLX, SSQ)	22 pengguna
[23]	2022	Perbandingan aplikasi <i>mind mapping</i> : <i>Xmind</i> dan <i>MindMaster</i>	<i>Task performance</i> (time, error rate), eye-tracking, pre-test kuesioner, post-test interview	20 pengguna
[20]	2023	Perbandingan AR vs VR untuk pembuatan <i>bounding box</i> objek nyata	<i>Task performance</i> (time, IoU), kuesioner	12 pengguna
[25]	2023	Evaluasi UX pada aplikasi pembelajaran Moodle dan Open edX	Kuesioner (SUS, AttrakDiff) performa tugas (time, success rate), physiological data (skin conductance, heart rate)	15 pengguna
[15]	2025	Perbandingan metode <i>tree testing</i> (<i>tree-visible</i> , <i>path-visible</i> , dan <i>compact</i>) untuk evaluasi representasi UI dalam arsitektur informasi	<i>Tree testing</i> (8 implisit dan 2 eksplisit feedback), kuesioner	180 pengguna

Publ ikasi	Tah un	Konteks Evaluasi	Teknik Pengumpulan Data	Partisipa n
[21]	2025	Perbandingan <i>spreadsheet</i> programming dengan dan tanpa <i>sheet-defined</i> <i>functional abstractions</i> (lambdas) pada aplikasi Lattice	<i>Task performance</i> (time, error and success rate), kuesioner (UEX), wawancara	24 pengguna

Berdasarkan Tabel 4, teknik pengumpulan data yang paling banyak digunakan adalah kombinasi antara task performance dan kuesioner. Pendekatan ini memungkinkan evaluasi dilakukan tidak hanya berdasarkan persepsi subjektif pengguna tetapi juga berdasarkan performa objektif selama menyelesaikan tugas tertentu.

Berdasarkan Tabel 4, teknik pengumpulan data yang digunakan dalam studi yang dianalisis dapat dikelompokkan ke dalam tiga bentuk utama. Pertama, studi yang menggunakan kuesioner sebagai teknik utama untuk menangkap persepsi pengguna. Kedua, sebagian besar studi mengombinasikan kuesioner dengan *task performance*, sehingga evaluasi tidak hanya didasarkan pada persepsi subjektif, tetapi juga pada performa pengguna seperti waktu penyelesaian tugas, tingkat keberhasilan, *error rate*, *precision*, atau metrik serupa. Ketiga, terdapat pendekatan khusus yang berfokus pada evaluasi struktur informasi, di mana peserta diminta menavigasi hierarki konten untuk menemukan lokasi yang tepat. Pendekatan ini berbeda dari teknik lain karena tidak mengevaluasi interaksi langsung dengan antarmuka, melainkan menilai kejelasan organisasi informasi berdasarkan tingkat keberhasilan dan waktu pencarian, yang kemudian dilengkapi dengan kuesioner untuk memperoleh persepsi pengguna [15].

Dari sisi partisipan, sebagian besar studi melaporkan jumlah partisipan dan karakteristik responden secara memadai, seperti mahasiswa, pengguna umum, pengguna pembelajar atau ahli, serta pengguna lansia. Namun demikian, beberapa studi hanya menjelaskan karakteristik partisipan secara umum sehingga informasi demografis atau latar belakang pengguna tidak selalu tersedia secara rinci. Variasi kelengkapan pelaporan ini juga tercermin pada hasil *Quality Assessment* yang menunjukkan adanya perbedaan tingkat detail pelaporan antar studi. Jumlah partisipan yang terlibat dalam penelitian bervariasi antara 10 hingga 180 orang. Adapun konteks evaluasi yang digunakan juga beragam, mencakup evaluasi antarmuka aplikasi, *software* tertentu, *Virtual Reality* (VR), *Augmented Reality* (AR), platform pembelajaran daring, hingga evaluasi arsitektur informasi. Variasi tersebut menunjukkan bahwa evaluasi UI/UX diterapkan pada berbagai domain sistem dengan kebutuhan, karakteristik pengguna, dan konteks penggunaan yang berbeda.

C. Aspek Evaluasi

Aspek UI/UX yang dievaluasi dalam studi yang dianalisis menunjukkan variasi dimensi yang digunakan untuk menilai kualitas antarmuka dan pengalaman pengguna.

TABEL 5. ASPEK YANG DIUKUR DALAM STUDI KOMPARASI EVALUASI UI/UX

Aspek Evaluasi	Indikator Umum	Paper
Efektivitas (<i>Effectiveness</i>)	Tingkat keberhasilan penyelesaian tugas, akurasi	[7], [16]–[18], [21], [23]–[25]

	interaksi, kesesuaian hasil dengan tujuan pengguna.	
Efisiensi (<i>Efficiency</i>)	Waktu penyelesaian tugas, jumlah langkah atau usaha yang diperlukan pengguna.	[5], [7], [16]–[21], [23]–[25]
Kepuasan pengguna (<i>User satisfaction</i>)	Persepsi kepuasan, kenyamanan, dan kemudahan penggunaan yang dilaporkan pengguna.	[7], [16], [19] [23], [24]
Aspek UX (<i>non-fungsional</i>)	<i>Attractiveness</i> , perspicuity (kejelasan), <i>stimulation</i> , <i>novelty</i> , dan pengalaman emosional.	[5], [25], [21], [23], [15]
Beban penggunaan (<i>workload / discomfort</i>)	Kelelahan, <i>motion sickness</i> , atau beban kognitif selama penggunaan sistem.	[[18]–[22], [24], [25]
Respons fisiologis / emosi	<i>Skin conductance</i> , <i>heart rate</i> , stres pengguna	[23]–[25]
Struktur informasi / <i>findability</i>	Navigasi, struktur hierarki, path deviation	[15]

Berdasarkan hasil ekstraksi data yang dirangkum pada Tabel 5, aspek evaluasi dapat dikelompokkan ke dalam beberapa kategori utama, yaitu efektivitas, efisiensi, kepuasan pengguna, aspek UX (*non-fungsional*), beban penggunaan, respons fisiologis, serta struktur informasi. Variasi aspek ini menunjukkan bahwa evaluasi tidak hanya berfokus pada performa tugas, tetapi juga mencakup persepsi pengguna, pengalaman emosional, serta faktor kognitif dan fisiologis selama interaksi dengan sistem.

Berdasarkan Tabel 5, aspek evaluasi yang paling dominan digunakan dalam studi komparatif UI/UX adalah efektivitas dan efisiensi. Kedua aspek ini umumnya diukur melalui metrik performa tugas seperti tingkat keberhasilan, akurasi, serta waktu penyelesaian tugas. Kepuasan pengguna juga sering dilaporkan, namun sebagian besar diperoleh melalui instrumen kuesioner sehingga bersifat subjektif. Dominasi ketiga aspek tersebut menunjukkan bahwa evaluasi dalam studi komparatif masih berfokus pada performa interaksi pengguna dan persepsi terhadap kemudahan penggunaan sistem.

Selain aspek utama tersebut, beberapa studi menambahkan dimensi UX non-fungsional, beban penggunaan, serta respons fisiologis sebagai pelengkap evaluasi. Dimensi UX seperti *attractiveness* dan *stimulation* digunakan untuk menangkap pengalaman emosional pengguna, sedangkan beban penggunaan menilai tingkat kelelahan atau beban kognitif selama interaksi. Penggunaan respons fisiologis dan evaluasi struktur informasi hanya muncul pada sebagian kecil studi, yang menunjukkan bahwa indikator tersebut masih terbatas penggunaannya. Secara keseluruhan, hasil ini menunjukkan bahwa evaluasi UI/UX dalam studi komparatif cenderung mengombinasikan metrik performa dan persepsi pengguna, sementara indikator tambahan digunakan untuk memperkaya analisis.

D. Hasil Perbandingan Berdasarkan Metode Objek yang Dibandingkan

Hasil perbandingan dalam studi komparatif UI/UX menunjukkan pola temuan yang berbeda bergantung pada objek yang dibandingkan. Berdasarkan pengelompokan pada RQ1, sintesis hasil perbandingan disajikan dalam tiga kategori utama, yaitu perbandingan metode evaluasi, perbandingan teknik atau pendekatan interaksi, serta perbandingan software atau sistem yang dievaluasi. Pengelompokan ini digunakan untuk mengidentifikasi pola

temuan yang lebih terstruktur sesuai konteks evaluasi yang dilakukan pada masing-masing studi.

- Perbandingan metode evaluasi UI/UX

Perbandingan metode evaluasi menunjukkan bahwa perbedaan instrumen pengukuran menghasilkan luaran yang tidak konsisten meskipun diterapkan pada sistem dan partisipan yang sama. Studi oleh Nakamura et al. membandingkan dua metode evaluasi UX, yaitu *AttrakDiff* dan *User Experience Questionnaire* (UEQ). Hasil menunjukkan bahwa dimensi pragmatik dan hedonic pada UEQ mengalami perubahan yang berbeda dibandingkan skor *AttrakDiff*. Sementara itu metode *Sentence Completion* mengungkap isu penggunaan yang tidak terdeteksi oleh kedua instrumen kuantitatif [5]. Ketidaksesuaian ini menyebabkan perbedaan interpretasi terhadap kualitas pengalaman pengguna dan berdampak langsung pada rekomendasi pengembangan sistem, yang menegaskan bahwa masing-masing metode memiliki sensitivitas berbeda terhadap aspek evaluasi yang diukur.

Perbedaan hasil juga terlihat pada studi yang membandingkan varian metode *tree testing*, yaitu *tree-visible*, *path-visible*, dan *compact*, dalam evaluasi arsitektur informasi [15]. Varian *tree-visible*, yang menampilkan struktur navigasi secara penuh, menghasilkan tingkat keberhasilan tugas yang lebih tinggi dan navigasi yang lebih konsisten dibandingkan varian lain. Sebaliknya, varian *compact* meningkatkan kesalahan navigasi dan waktu pencarian karena pengguna harus mengingat konteks struktur secara mental [15]. Perbedaan tersebut menunjukkan bahwa variasi implementasi metode evaluasi dapat memengaruhi metrik objektif seperti *task success rate* dan *completion time*, meskipun objek arsitektur informasi yang dievaluasi sama.

Secara keseluruhan, hasil perbandingan menunjukkan bahwa karakteristik metode evaluasi memengaruhi luaran evaluasi yang diperoleh. Penggunaan metode yang berbeda dapat saling melengkapi karena masing-masing menekankan dimensi pengalaman yang berbeda. Selain itu, pemilihan metode evaluasi sebaiknya disesuaikan dengan tujuan evaluasi dan aspek evaluasi yang ingin diukur.

- Perbandingan teknik/interaksi

Studi yang membandingkan teknik atau pendekatan interaksi menunjukkan bahwa berbagai metrik digunakan untuk mengevaluasi suatu sistem, terutama untuk mengukur efisiensi dan efektivitas. Sebagian besar penelitian dalam kategori ini menggunakan metrik performa seperti waktu penyelesaian tugas, tingkat keberhasilan, serta jumlah kesalahan untuk mengidentifikasi perbedaan antar teknik interaksi. Hasil perbandingan menunjukkan bahwa pendekatan interaksi yang berbeda dapat menghasilkan variasi signifikan pada efisiensi dan efektivitas, terutama pada sistem dengan kompleksitas interaksi tinggi seperti VR dan AR [20].

Beberapa studi menunjukkan bahwa teknik interaksi berbasis *controller* menghasilkan performa yang lebih stabil dibandingkan *gesture-based interaction*, terutama dalam tugas manipulasi objek dan navigasi ruang [22]. Selain itu, pendekatan interaksi multimodal juga dilaporkan dapat meningkatkan efisiensi dibandingkan pendekatan unimodal karena memberikan fleksibilitas dalam memilih metode input. Namun, peningkatan performa tersebut tidak selalu diikuti oleh peningkatan kepuasan pengguna, sehingga hasil evaluasi bergantung pada dimensi evaluasi yang diukur. Temuan ini menunjukkan bahwa teknik interaksi yang lebih kompleks atau lebih natural tidak selalu menghasilkan

performa yang lebih baik. Pada beberapa kasus, interaksi yang lebih sederhana atau konvensional justru menghasilkan waktu penyelesaian tugas yang lebih cepat dan kesalahan yang lebih sedikit.

Hasil perbandingan teknik interaksi juga menunjukkan bahwa metrik performa lebih sensitif dalam mengidentifikasi perbedaan antar teknik, sedangkan persepsi pengguna cenderung dipengaruhi oleh faktor pengalaman dan kenyamanan. Oleh karena itu, evaluasi teknik interaksi sebaiknya mempertimbangkan kombinasi metrik objektif dan subjektif untuk memperoleh gambaran performa UI/UX yang lebih menyeluruh.

- Perbandingan performa software

Studi yang membandingkan performa *software* menunjukkan bahwa evaluasi digunakan untuk mengidentifikasi kelebihan dan keterbatasan sistem yang memiliki fungsi serupa [23]-[25]. Hasil menunjukkan bahwa perbedaan desain dan fitur navigasi dapat memengaruhi efisiensi penggunaan, meskipun fungsi utama sistem yang dibandingkan relatif sama.

Beberapa penelitian menemukan bahwa *software* dengan struktur informasi yang lebih jelas dan navigasi yang konsisten menghasilkan tingkat keberhasilan tugas yang lebih tinggi dan kesalahan yang lebih sedikit [15], [23]-[25]. Namun demikian, platform yang lebih kompleks dengan fitur tambahan terkadang memperoleh penilaian pengalaman pengguna yang lebih tinggi meskipun performa tugasnya tidak selalu lebih baik. Hal ini menunjukkan bahwa persepsi evaluasi UI/UX tidak hanya dipengaruhi oleh efisiensi, tetapi juga oleh faktor pengalaman dan fleksibilitas penggunaan.

Selain itu, hasil perbandingan antar platform menunjukkan bahwa tidak ada sistem yang secara konsisten unggul pada seluruh dimensi. Sebuah platform dapat lebih efisien dalam penyelesaian tugas, sementara platform lain lebih unggul pada kepuasan pengguna atau kemudahan belajar. Temuan ini menunjukkan bahwa evaluasi UI/UX pada perbandingan *software* perlu mempertimbangkan berbagai dimensi secara bersamaan agar dapat memberikan gambaran yang lebih komprehensif terhadap kualitas sistem yang dibandingkan.

Hasil sintesis pada tiga kategori menunjukkan bahwa perbandingan metode evaluasi menekankan perbedaan sensitivitas instrumen, perbandingan teknik interaksi menunjukkan variasi performa dan pengalaman pengguna, sedangkan perbandingan performa *software* mengungkap kelebihan dan keterbatasan masing-masing sistem pada dimensi evaluasi yang berbeda. Rangkuman temuan dari ketiga kategori tersebut disajikan secara visual pada Gambar 3.



Gambar 3. Sintesis hasil perbandingan pendekatan evaluasi dalam studi komparatif UI/UX

E. Diskusi

Hasil analisis menunjukkan bahwa evaluasi dalam studi komparatif tidak hanya berfokus pada *usability*, tetapi mencakup kerangka evaluasi UI/UX yang lebih luas. Menurut ISO 9241-210:2019, evaluasi UX akan menghasilkan penilaian akan persepsi dan respons pengguna terhadap suatu sistem, yang mencakup aspek emosional, kognitif, dan perilaku sebelum, selama, dan setelah interaksi. Evaluasi UX dalam kerangka ini tidak hanya dipengaruhi oleh karakteristik sistem, tetapi juga oleh kondisi pengguna serta konteks penggunaan, sehingga evaluasi bersifat multidimensi [6]. Perbedaan pendekatan evaluasi, teknik interaksi, dan karakteristik sistem yang dibandingkan menyebabkan variasi hasil yang tidak hanya mencerminkan kualitas sistem, tetapi juga sensitivitas metode terhadap aspek tertentu. Hal ini menunjukkan bahwa hasil evaluasi UI/UX bersifat kontekstual dan dipengaruhi oleh kombinasi metode serta metrik yang digunakan.

Implikasi pertama adalah pentingnya penggunaan pendekatan evaluasi multi-metode dalam studi UI/UX. Metrik berbasis performa seperti waktu penyelesaian tugas dan tingkat keberhasilan lebih sensitif dalam membedakan efisiensi antar alternatif desain, sementara metode subjektif seperti kuesioner mampu menangkap persepsi, kenyamanan, dan pengalaman pengguna. Selain itu, beberapa studi juga menunjukkan bahwa data tambahan seperti *eye-tracking* dan respons fisiologis dapat mengungkap aspek interaksi yang tidak terdeteksi melalui metrik tradisional. Oleh karena itu, integrasi metrik kuantitatif, temuan kualitatif, dan data perilaku objektif memberikan gambaran holistik mengenai efektivitas suatu sistem [26].

Implikasi kedua berkaitan dengan hubungan antara teknik interaksi dan hasil evaluasi UI/UX. Studi menunjukkan bahwa teknik interaksi lebih baru atau lebih kompleks tidak selalu menghasilkan performa yang lebih baik. Dalam beberapa kasus, pendekatan konvensional justru memberikan efisiensi yang lebih tinggi dan tingkat kesalahan yang lebih rendah. Hal ini mengindikasikan adanya *trade-off* antara efisiensi dan pengalaman pengguna, di mana peningkatan teknologi atau teknik interaksi dapat meningkatkan kenyamanan atau *engagement*, tetapi tidak selalu sejalan dengan performa tugas [19].

Implikasi ketiga adalah pentingnya pendekatan multidimensi dalam perbandingan performa *software*. Tidak adanya sistem yang secara konsisten unggul pada seluruh dimensi menunjukkan bahwa evaluasi UI/UX tidak dapat bergantung pada satu indikator tunggal [27]. Performa tugas,

persepsi pengguna, serta aspek pengalaman seperti kenyamanan dan fleksibilitas perlu dipertimbangkan secara bersamaan untuk memahami kelebihan dan keterbatasan masing-masing sistem. Pendekatan ini memungkinkan identifikasi *trade-off* antar dimensi dan mendukung pengambilan keputusan desain yang lebih tepat.

Secara keseluruhan, temuan studi ini menegaskan bahwa evaluasi UI/UX dalam konteks komparatif bersifat multidimensi dan kontekstual. Oleh karena itu, perancangan evaluasi perlu dilakukan secara sistematis dengan mempertimbangkan pemilihan metode, teknik pengumpulan data, jenis metrik, serta kesesuaiannya dengan karakteristik sistem dan dimensi pengalaman pengguna yang menjadi fokus analisis. Temuan ini memperkuat pentingnya pemilihan pendekatan evaluasi yang sesuai dengan tujuan dan konteks penggunaan sistem. Namun demikian, studi ini juga mengidentifikasi adanya kesenjangan penelitian, khususnya pada masih terbatasnya jumlah studi yang secara eksplisit membandingkan metode evaluasi UI/UX. Kesenjangan ini menunjukkan perlunya penelitian lebih lanjut untuk mengembangkan kerangka evaluasi yang lebih terintegrasi, sistematis, dan komprehensif dalam studi komparatif UI/UX.

Walaupun studi ini telah dapat menjawab jenis pendekatan evaluasi UI/UX yang digunakan dalam studi komparatif yang dipelajari, konteks evaluasi, teknik pengumpulan data, serta dimensi yang digunakan dalam menilai kualitas UI/UX dan hasil perbandingan yang dilakukan, studi ini masih memiliki beberapa keterbatasan. Keterbatasan tersebut antara lain jumlah artikel yang dianalisis relatif terbatas, ruang lingkup pencarian yang hanya mencakup satu basis data, serta pemilihan artikel yang bergantung pada ketersediaan akses *full text*. Selain itu, hasil sintesis yang disajikan masih dipengaruhi oleh heterogenitas metode evaluasi, karakteristik partisipan, konteks aplikasi, serta variasi tingkat kelengkapan pelaporan pada masing-masing studi, sehingga generalisasi temuan perlu dilakukan secara hati-hati. Oleh karena itu, penelitian selanjutnya dapat memperluas cakupan sumber literatur, menambah jumlah studi yang dianalisis, serta mengkaji pendekatan evaluasi UI/UX pada domain aplikasi yang lebih beragam untuk memperoleh pemahaman yang lebih komprehensif.

IV. KESIMPULAN

Penelitian ini melakukan *Systematic Literature Review* terhadap 13 artikel yang membahas evaluasi UI/UX dalam studi komparatif yang dipublikasikan pada periode 2020–2025. Hasil penelitian menunjukkan bahwa objek yang dibandingkan dalam studi komparatif UI/UX dapat dikelompokkan ke dalam tiga kategori utama, yaitu perbandingan metode evaluasi, perbandingan teknik atau pendekatan interaksi, serta perbandingan *software* atau sistem yang dievaluasi. Sebagian besar penelitian menggunakan evaluasi UI/UX sebagai alat untuk membandingkan teknik interaksi (61,5%) dan performa *software* (23,1%), sedangkan studi yang secara langsung membandingkan metode evaluasi masih relatif terbatas (15,4%). Dari sisi prosedur evaluasi, sebagian besar penelitian menggunakan kombinasi kuesioner dan task performance untuk mengukur aspek subjektif maupun objektif dari interaksi pengguna. Karakteristik partisipan dan konteks evaluasi yang beragam menunjukkan bahwa evaluasi UI/UX diterapkan pada berbagai domain sistem, mulai dari aplikasi umum, perangkat lunak khusus, *Virtual Reality*

(VR), *Augmented Reality* (AR), hingga platform pembelajaran daring dan arsitektur informasi. Hasil sintesis menunjukkan bahwa aspek evaluasi yang paling sering digunakan adalah efektivitas, efisiensi, dan kepuasan pengguna. Selain itu, beberapa studi juga mengukur aspek pengalaman pengguna yang lebih luas, seperti *attractiveness*, *stimulation*, beban penggunaan, respons fisiologis, dan struktur informasi. Temuan ini menunjukkan bahwa evaluasi UI/UX dalam studi komparatif tidak hanya berfokus pada *usability*, tetapi juga mencakup dimensi pengalaman pengguna yang lebih komprehensif. Hasil perbandingan menunjukkan bahwa metode evaluasi yang berbeda dapat menghasilkan interpretasi yang berbeda terhadap kualitas pengalaman pengguna. Pada perbandingan teknik interaksi, metrik performa terbukti lebih sensitif dalam mengidentifikasi perbedaan antar teknik dibandingkan persepsi pengguna. Sementara itu, pada perbandingan software, tidak ditemukan sistem yang secara konsisten unggul pada seluruh dimensi evaluasi, sehingga pendekatan multidimensi diperlukan untuk memperoleh gambaran kualitas sistem yang lebih menyeluruh. Selain itu, proses *Quality Assessment* menunjukkan bahwa seluruh studi yang disintesis memiliki kualitas metodologis yang memadai sehingga layak digunakan dalam proses ekstraksi data dan sintesis temuan. Hal ini meningkatkan kredibilitas dan validitas hasil penelitian dalam menggambarkan karakteristik evaluasi UI/UX pada studi komparatif. Kontribusi utama penelitian ini adalah menyediakan pemetaan sistematis mengenai objek yang dibandingkan, prosedur evaluasi, aspek UI/UX yang diukur, serta hasil perbandingan yang dilaporkan dalam studi komparatif UI/UX. Penelitian ini juga mengidentifikasi masih terbatasnya studi yang secara langsung membandingkan metode evaluasi UI/UX, sehingga membuka peluang penelitian lanjutan untuk mengkaji validitas, sensitivitas, konsistensi, serta kesesuaian berbagai metode evaluasi pada konteks penggunaan yang berbeda. Penelitian ini masih memiliki beberapa keterbatasan, antara lain jumlah artikel yang relatif terbatas, penggunaan satu basis data utama, serta heterogenitas metode evaluasi, karakteristik partisipan, dan konteks aplikasi yang digunakan pada masing-masing studi. Oleh karena itu, penelitian selanjutnya dapat memperluas cakupan sumber literatur, meningkatkan jumlah studi yang dianalisis, serta mengkaji perbandingan metode evaluasi UI/UX secara lebih mendalam untuk mendukung pengembangan kerangka evaluasi yang lebih sistematis dan komprehensif.

REFERENCES

- [1] J. Feng and B. Yu, "Key Factors Influencing Educational Technology Adoption in Higher Education: A Systematic Review," *PLOS Digital Health*, vol. 4, no. 4, Art. no. e0000764, 2025, doi: 10.1371/journal.pdig.0000764.
- [2] L. Lemma and C. Ma, "The Role of Information Systems in Business Firms Competitiveness: Integrated Review Paper from Business Perspective," *International Research Journal of Natural Science and Technology*, vol. 2, no. 4, pp. 29–42, 2020.
- [3] A. I. Stoumpos and F. Kitsios, "Digital Transformation in Healthcare: Technology Acceptance and Its Applications," *International Journal of Environmental Research and Public Health*, vol. 20, no. 4, Art. no. 3407, 2023, doi: 10.3390/ijerph20043407.
- [4] B. Nurohman and A. Aziz, "Analisis UI/UX terhadap Perancangan Toko Online dengan Menggunakan Pendekatan Human Centered Design," *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer*, vol. 10, no. 1, pp. 27–37, 2024.
- [5] W. T. Nakamura, I. Ahmed, D. Redmiles, E. Oliveira, D. Fernandes, E. H. T. de Oliveira, and T. Conte, "Are UX Evaluation Methods Providing the Same Big Picture?" *Sensors*, vol. 21, no. 10, Art. no. 3480, 2021, doi: 10.3390/s21103480.
- [6] S. Ntoa, "Usability and User Experience Evaluation in Intelligent Environments: A Review and Reappraisal," *International Journal of Human-Computer Interaction**, pp. 1–30, 2024, doi: 10.1080/10447318.2024.2394724.
- [7] L. T. De Paolis and V. De Luca, "The Effects of Touchless Interaction on Usability and Sense of Presence in a Virtual Environment," *Virtual Reality*, vol. 26, no. 4, pp. 1551–1571, 2022, doi: 10.1007/s10055-022-00647-1.
- [8] P. Theodorou, K. Tsiligkos, and A. Meliones, "Multi-Sensor Data Fusion Solutions for Blind and Visually Impaired: Research and Commercial Navigation Applications for Indoor and Outdoor Spaces," *Sensors*, vol. 23, no. 12, Art. no. 5411, 2023, doi: 10.3390/s23125411.
- [9] M. Sandesara, U. Bodkhe, S. Tanwar, N. Kumar, M. Alazab, and P. R. Kshirsagar, "Design and Experience of Mobile Applications: A Pilot Survey," *Mathematics*, vol. 10, no. 14, Art. no. 2380, 2022, doi: 10.3390/math10142380.
- [10] P. Hu, B. Gao, K. Li, and B. Gao, "Learning Interpreting in Virtual Reality: A Scoping Review," *Interactive Learning Environments*, vol. 33, no. 1, pp. 347–366, 2025, doi: 10.1080/10494820.2024.2347304.
- [11] M. J. Page et al., "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews," *BMJ*, vol. 372, Art. no. n71, 2021, doi: 10.1136/bmj.n71.
- [12] B. Kitchenham and S. Charters, *Guidelines for Performing Systematic Literature Reviews in Software Engineering*, EBSE Technical Report EBSE-2007-01, Keele University and Durham University, 2007.
- [13] B. Kitchenham, L. Madeyski, and D. Budgen, "SEGRESS: Software Engineering Guidelines for Reporting Secondary Studies," *IEEE Transactions on Software Engineering*, vol. 49, no. 3, pp. 1273–1298, Mar. 2023, doi: 10.1109/TSE.2022.3174092.
- [14] N. Nyoni, "A Narrative Guideline to Structure the Analysis Stage of Systematic Literature Reviews," *Electronic Journal of Business Research Methods*, vol. 23, no. 1, pp. 1–15, 2025.
- [15] E. Kuric, "Validation of Information Architecture: Cross-Methodological Comparison of Tree Testing Variants and Prototype User Testing," *Information and Software Technology*, vol. 183, p. 107740, 2025, doi: 10.1016/j.infsof.2025.107740.
- [16] N. A. Ismail, N. A. A. Majid, N. H. A. Wahab, and F. Mohamed, "A Comparative Study of Unimodal and

- Multimodal Interactions for Digital TV Remote Control Mobile Application among Elderly,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, no. 7, pp. 157–165, 2021, doi: 10.14569/IJACSA.2021.0120718.
- [17] M. Fechter, B. Schleich, and S. Wartzack, “Comparative Evaluation of WIMP and Immersive Natural Finger Interaction: A User Study on CAD Assembly Modeling,” *Virtual Reality*, vol. 26, no. 1, pp. 143–158, 2022, doi: 10.1007/s10055-021-00543-0.
- [18] M. Allgaier et al., “A Comparison of Input Devices for Precise Interaction Tasks in VR-Based Surgical Planning and Training,” *Computers in Biology and Medicine*, vol. 145, p. 105429, 2022, doi: 10.1016/j.compbiomed.2022.105429.
- [19] M. Di Vincenzo, F. Palini, M. De Marsico, A. M. Borghi, and G. Baldassarre, “A Natural Human-Drone Embodied Interface: Empirical Comparison with a Traditional Interface,” *Frontiers in Neurorobotics*, vol. 16, 2022, doi: 10.3389/fnbot.2022.898859.
- [20] N. Kyaw, M. Gu, E. Croft, and A. Cosgun, “Comparing Usability of Augmented Reality and Virtual Reality for Creating Virtual Bounding Boxes of Real Objects,” *Applied Sciences*, 2023.
- [21] A. Kudebayeva, C. Harte, N. Bryan-Kinns, and T. Stockman, “A Comparative Study Evaluating the User Experience of Spreadsheet Programming with Sheet-Defined Functional Abstractions,” *International Journal of Human-Computer Interaction*, vol. 41, no. 9, pp. 5445–5461, 2024, doi: 10.1080/10447318.2024.2361209.
- [22] F. Buttussi and L. Chittaro, “Locomotion in Place in Virtual Reality: A Comparative Evaluation of Joystick, Teleport, and Leaning,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 1, pp. 125–136, 2021, doi: 10.1109/TVCG.2019.2928304.
- [23] J. Wang, Z. Xu, X. Wang, and J. Lu, “A Comparative Research on Usability and User Experience of User Interface Design Software,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 13, no. 8, pp. 21–29, 2022, doi: 10.14569/IJACSA.2022.0130804.
- [24] J. Wang, X. Wang, J. Lu, and Z. Xu, “Investigating the User Experience of Mind Map Software: A Comparative Study Based on Eye Tracking,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 13, no. 11, pp. 1–11, 2022, doi: 10.14569/IJACSA.2022.0131101.
- [25] A. Liapis, V. Maratou, and T. Panagiotakopoulos, “UX Evaluation of Open MOOC Platforms: A Comparative Study Between Moodle and Open edX Combining User Interaction Metrics and Wearable Biosensors,” *Interactive Learning Environments*, 2023, doi: 10.1080/10494820.2022.2048674.
- [26] P. Tewari, “UX Principles for Modern UI/UX Design and Their Measurement: A Framework for Digital Product Excellence,” *Journal of Computer Science and Technology Studies*, vol. 7, pp. 898–907, 2025.
- [27] A. Berni and Y. Borgianni, “From the Definition of User Experience to a Framework to Classify ITS Applications in Design,” in *Proceedings of the 23rd International Conference on Engineering Design (ICED21)*, Gothenburg, Sweden, Aug. 2021, pp. 1627–1636, doi: 10.1017/pds.2021.424.