

Evaluating Interactive Assessment Features of an Web-Based Learning System Using UEQ

Eka Larasati Amalia^{1*}, Dimas Wahyu Wibowo², Venny Meida Hersianty³, Naufal Yukafi Ridlo⁴,
Mustika Mentari⁵

Abstract—Object-Oriented Programming (OOP) is one of the most conceptually demanding subjects in vocational software engineering education. However, students are usually assessed with multiple-choice items that test recognition rather than their ability to construct code. Fill-in-the-blank and radio-button formats address this gap in complementary ways. However, their combined use in a single vocational OOP platform has not been evaluated from a user-experience perspective. The purpose of this study is to evaluate the user experience of OOPedia. This web-based OOP learning system integrates both formats and positions the results against the UEQ international benchmark. The study used a two-stage method: OOPedia was first developed using a prototype development methodology, and its user experience was measured using the User Experience Questionnaire (UEQ) across six dimensions. Participants were 32 Software Engineering (SE) students, aged 16–17, taking the OOP course at a vocational high school in Malang, Indonesia, all of whom had used both features beforehand. Dimension means were tested against the neutral midpoint using one-sample t-tests, compared with the UEQ global benchmark, and assessed for reliability using Cronbach's α . The results show that all six dimensions scored significantly above neutral ($p < 0.001$) and met the benchmark: Stimulation and Attractiveness reached the Excellent band, and the other four reached Good or Above Average. In contrast, novelty showed low reliability ($\alpha = 0.52$) and was reported with caution. The practical impact is an empirical UX baseline that designers of school e-learning platforms can use when selecting interactive assessment formats, ahead of the controlled learning-outcome study recommended as future work.

Index Terms—User experience questionnaire, web-based learning system, interactive assessment, fill-in-the-blank, radio-button.

I. INTRODUCTION

The shift toward web-based learning in Indonesian vocational education has made assessment design a critical variable in whether students actually develop the competencies they are taught, particularly as personalized and adaptive learning approaches become increasingly prominent in digital education [1], [2]. Digital platforms now support both content delivery and skill practice in SE programs [3], [4], [5], yet most e-learning systems in this context still rely on multiple-choice questions that test recognition rather than the ability to produce OOP code [6]. This condition matters because OOP is unusually hard to learn: failure rates in introductory programming average 28–33% across institutions worldwide [7], [8], and documented difficulties with inheritance, encapsulation, and the class concept persist even among students who pass [9]. In asynchronous settings, where immediate teacher feedback is unavailable. This mismatch between recognition-based assessment and construction-based learning is sharpened further [10]. Vocational students at Indonesian school face this problem acutely: they are expected to write working OOP code, but the assessments they encounter primarily ask them to recognize it.

Two assessment formats address this mismatch in complementary ways. Fill-in-the-blank items require students to retrieve and produce OOP syntax from memory, process research links to deeper cognitive processing and stronger retention than multiple-choice selection [6]. Radio-button items, when paired with instant feedback, support efficient formative checks of basic conceptual understanding at lower cognitive cost [6]. Combining both within a single platform can enable retrieval and recognition within a single session. Each format has been studied separately in programming education contexts [11]. What has not been examined is whether combining them in a single vocational OOP system produces a user experience that is acceptable, motivating, and internationally benchmark competitive.

This study evaluates the user experience of OOPedia, a web-based OOP learning system that integrates both

Received: 15 May 2026; Revised: 16 June 2026; Accepted: 24 June 2026.

*Corresponding author

¹Eka Larasati Amalia, Department of Information Technology, Politeknik Negeri Malang, Malang, Indonesia (e-mail: eka.larasati@polinema.ac.id).

²Dimas Wahyu Wibowo, Department of Information Technology, Politeknik Negeri Malang, Malang, Indonesia (e-mail: dimas.w@polinema.ac.id).

³Venny Meida Hersianty, Department of Information Technology, Politeknik Negeri Malang, Malang, Indonesia (e-mail: vnymeida@gmail.com).

⁴Naufal Yukafi Ridlo, Department of Information Technology, Politeknik Negeri Malang, Malang, Indonesia (e-mail: yukafiwork@gmail.com).

⁵Mustika Mentari, Department of Information and Communication Systems, Okayama University, Okayama, Japan (e-mail: pqt85hm5@s.okayama-u.ac.jp).

fill-in-the-blank and radio-button assessment features, using UEQ [12]. The UEQ measures six dimensions (attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty) and provides a global benchmark dataset that allows findings to be positioned within an international rather than local frame [13], [14]. The evaluation was conducted with year-11 Software Engineering students at a vocational high school. The population needs most effective OOP assessment design and is least represented in published UEQ studies. Without evidence on how vocational students experience combined interactive assessment, designers of school e-learning platforms lack an empirical basis for choosing among assessment format options.

This study makes two contributions. First, to our knowledge, no prior study has applied a UEQ-benchmarked evaluation to the combined use of construction-based and recognition-based interactive assessment in a vocational OOP platform, making this the first study to position such a design against an international UX benchmark. Second, it provides a UX profile from a TVET cohort, a population underrepresented in published UEQ research. The three research questions are:

- **RQ1.** How do vocational students perceive the user experience of combined fill-in-the-blank and radio-button assessment features in OOPedia across the six UEQ dimensions?
- **RQ2.** How does the user experience of OOPedia's assessment features compare against the UEQ global benchmark intervals?
- **RQ3.** Which UEQ dimensions indicate the strongest signal for design improvement, and what specific design directions do those signals suggest?

The authors note explicitly that the present study evaluates *user experience*, not *learning efficacy*. Learning outcomes are not measured here; a controlled comparative study is recommended as future work and is discussed in Section VI. Section II reviews related work and identifies the research gap. Section III describes OOPedia and the evaluation method. Section IV reports results. Section V discusses the findings. Section VI concludes.

II. RELATED WORK

A. OOP Learning Difficulties in Vocational and Online Contexts

OOP is consistently identified as one of the hardest programming paradigms to learn at the introductory level. The core difficulty is not syntactic but conceptual: students struggle to shift from procedural thinking to object-oriented abstraction, particularly with respect to class hierarchies, encapsulation, and inheritance [15]. Gutiérrez et al.'s systematic review of 56 studies ranked these as the dominant recurring obstacles across teaching contexts [15]. In contrast, another study showed that procedural mental models persist even in students who have already passed earlier programming courses [9]. A prior study conducted a twelve-year survey across 161 institutions found an average failure rate of 28% in introductory programming [7]; meanwhile, the independent analysis of CS1 data put the figure

at approximately 33% [16]. These are not outliers; they represent a stable, cross-institutional pattern.

The problem is compounded in online and asynchronous settings, where the absence of immediate teacher feedback removes a key scaffold for students still building their OOP vocabulary [10]. The immersive interactive media can improve OOP understanding at the university level [17], but their study does not address asynchronous vocational contexts where infrastructure and access vary significantly [4]. In Indonesian SMK specifically, Habibi et al. found that access to digital technology is geographically uneven, meaning instructional design choices, including assessment format, carry disproportionate weight in determining learning outcomes [4]. Others demonstrated that assessment design is not a peripheral variable but a primary determinant of whether e-learning systems are used [3].

The literature establishes that OOP is hard, that the difficulty is amplified in asynchronous vocational settings, and that assessment design directly shapes both cognitive engagement and platform adoption. What it does not address is which specific assessment formats best serve this population or how students in this context experience those formats when delivered through a digital platform.

B. Assessment Format Effects on OOP Learning

Assessment format determines the cognitive operation students must perform: recognition (selecting from options) or construction (producing an answer). Multiple-choice-only assessment has been consistently associated with surface learning and recall, while production-based formats are linked to deeper processing and stronger transfer [6]. Meanwhile, a validation study in medical education showed that very-short-answer formats reduce cueing effects and prompt more authentic study behavior than multiple-choice, which is a finding with clear relevance for programming assessment, where code production is the target competency [6].

In programming education specifically, the fill-in-the-blank format closely maps onto authentic coding activity: completing an incomplete class definition or method header requires the learner to retrieve precise syntax and semantics, not merely recognize them [6], [18]. The radio-button format serves a different but complementary purpose, the efficient formative checks of conceptual understanding with low cognitive overhead, particularly effective when paired with immediate feedback [6]. A study that systematically mapped 195 publications found that even syntactically passing students often retain only shallow OOP design mastery, suggesting that neither format alone is sufficient [18].

The two formats have different but complementary cognitive profiles. Fill-in-the-blank probes for construction; radio buttons enable efficient formative recognition checks. Combining both within a single session could exercise both modes of OOP understanding. This fact has not been tested in a vocational OOP context, and no published study has evaluated how the combined design is perceived by users, which makes user experience evaluation the necessary next step before making learning-efficacy claims.

C. UEQ-based Evaluation of Educational Systems

The User Experience Questionnaire (UEQ), developed by Laugwitz, Held, and Schrepp [12], measures six dimensions across two quality groups, pragmatic (perspicuity, efficiency, dependability) and hedonic (stimulation, novelty), plus the overarching Attractiveness dimension. Another study published benchmark dataset, derived from hundreds of product evaluations, enables comparative positioning of a single study's results within a global reference frame [14]. This feature distinguishes the UEQ from instruments that produce only self-referential scores.

Several recent studies have applied the UEQ in Indonesian educational contexts. Then, [19] evaluated Zoom and Learn Quran Tajwid as online learning tools, demonstrating the instrument's cross-cultural reliability; however, their study focuses on content-delivery platforms rather than interactive assessment features. Reference [20] combined UEQ with usability testing of the Halodoc health application and found that UEQ alone missed interaction-level issues that later surfaced in usability testing, a methodological limitation worth noting for studies relying solely on the questionnaire. References [21] and [22] both reported novelty compression in Indonesian student samples, consistent with [14]'s observation that familiarity with comparable platforms deflates this dimension. Also, [23] found pragmatic dimensions to dominate hedonic ones in Moodle-based LMS evaluations, a pattern that the present study's results can be directly compared against.

UEQ has proven reliable and cross-culturally applicable across Indonesian educational platforms. However, all existing applications evaluate content-delivery systems, health applications, or general LMS tools; none target interactive assessment features embedded in a subject-specific OOP platform. Critically, the study does not benchmark a TVET cohort against the UEQ global dataset, leaving it in an unexamined methodological position.

D. Research Gap

The three subsections above converge on a specific, addressable gap. First, OOP learning difficulties in TVET contexts are well documented, but the evidence base for which assessment formats best serve this population remains thin [7], [9], [15]. Second, fill-in-the-blank and radio-button formats have been studied separately, but not evaluated in combination within a single OOP platform, nor from a user experience perspective [6], [18]. Third, the UEQ has been applied across Indonesian educational systems, but exclusively to content-delivery or general-purpose platforms; no study has used it to evaluate interactive assessment features in a vocational programming context, nor positioned such findings against the UEQ global benchmark [19], [22].

Methodologically, prior UEQ studies in this region used the instrument either as a standalone tool or alongside usability testing for full applications. The present study differs in two ways: it targets specific assessment *features* within a platform (not the platform as a whole), and it uses the UEQ benchmark comparison as the primary interpretive frame rather than a

supplementary one. This positions the study at the intersection of programming education research and UX evaluation as an intersection that, to our knowledge, no published work has occupied.

III. RESEARCH METHOD

This study evaluates the user experience of two interactive assessment features, fill-in-the-blank and radio-button, implemented in the OOPedia web-based OOP learning system. The research is organized in two phases: system development and system evaluation. Fig. 1 presents the overall research flow.

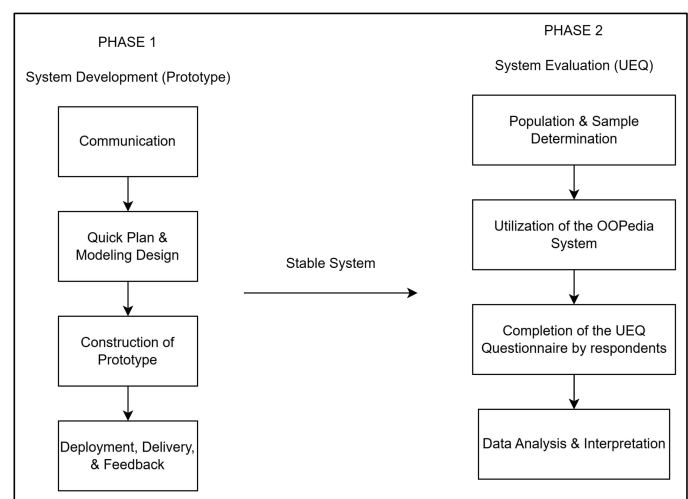


Fig. 1. Research flow.

A. System Development Phase

OOPedia was developed using the prototype methodology described by Pressman and Maxim [11], which proceeds through communication, quick plan, modeling quick design, construction of prototype, and deployment, delivery, and feedback. The communication step involved informal interviews and classroom observation with OOP teachers and software engineering students at the target vocational school, which surfaced two recurrent problems: passive, multiple-choice-only assessment and limited interactive practice focused on syntactic detail. The quick plan step scoped the system to two integrated assessment features intended to address those problems. The Modeling Quick Design step produced wireframes for question presentation, interaction, and feedback. The construction step implemented the two features using React.js on the front end and Express.js on the back end. The Deployment, Delivery, and Feedback step deployed the system in the classroom for an informal trial.

The development followed a single development-trial-refinement loop rather than multiple formally documented iteration cycles, and that the absence of formal iteration logs is a methodological limitation of this study. The single refinement cycle proceeded as follows: after the

initial prototype was deployed in a supervised classroom trial session, approximately one week before the main evaluation, two authors present in the classroom observed student interactions and noted recurring behavioral patterns. Two specific refinements were made in response to these observations: (1) the addition of instant feedback after radio-button submission, prompted by observing that students repeatedly attempted to re-submit items after receiving no acknowledgement; and (2) a session progress indicator showing the number of completed items, prompted by observing that students frequently asked the teacher how many questions remained, indicating they had lost track of their position in the assessment. Both refinements were implemented and verified before the formal UEQ evaluation session. The trial session involved the same Year-11 SE students who would later participate in the formal evaluation; all 32 attended. The session lasted approximately 45 minutes and was conducted under the direct supervision of two members of the research team, who took structured observational notes using a predefined checklist covering navigation errors, resubmission attempts, and verbal requests for clarification. The decision to proceed to formal evaluation, rather than conduct a second iteration, that was made when no new behavioral patterns emerged in the final 15 minutes of the trial and both features could be completed end-to-end without researcher intervention. The development phase ended when the two features were functionally stable and could be used end-to-end by students without supervision.

The two features are described in more detail below. The fill-in-the-blank feature presents an OOP question with one or more blank positions in a code snippet or concept statement, and the student types the missing fragment. The format requires active construction of OOP terminology and syntax and is intended to probe the student's ability to recall and produce, not only recognize, OOP elements [5]. The radio-button feature presents a multiple-choice question with one correct option and provides immediate feedback after submission, along with a progress indicator. The format is intended to support efficient formative checks of basic conceptual understanding [5]. Both features were implemented following user-centered design principles and shared a consistent interface style with the existing OOPedia platform. Fig. 2–4 shows the teacher and student views of the two features.

B. System Evaluation Phase

After the system was functionally stable, the user experience of the two assessment features was evaluated with the UEQ. The evaluation phase is described below in terms of the population and sample, the data collection procedure, the instrument, and the analysis.

1) Population and sample

The target population was year-11 SE students at a single vocational high school in Malang, East Java, Indonesia, enrolled in the OOP course during the 2024/2025 academic year. Participants ranged in age from 16 to 17 years. The sample comprised 32 students (18 male, 14 female). All participants had prior experience using web-based learning

platforms as part of their regular coursework, though none had used OOPedia's interactive assessment features before the study period.

Fig. 2. Teacher view: add multiple-choice (radio-button) question.

Fig. 3. Student view: multiple-choice (radio-button) question

Fig. 4. Student view: fill-in-the-blank question.

Participants were recruited by purposive sampling against three inclusion criteria: (a) currently enrolled in the OOP course, (b) had used both the fill-in-the-blank and the radio-button features in at least one OOPedia assessment session prior to the evaluation, and (c) provided written informed consent to participate. For participants under 18 years of age, written consent was also obtained from a parent or legal guardian prior to the study. The school principal provided institutional permission for the research to be conducted on school premises during regular school hours. No incentive was offered for participation. Students who chose not to participate faced no academic consequence and continued with alternative classroom activities during the evaluation session.

The rationale for purposive rather than random sampling is that user-experience evaluation requires that respondents have actually used the artifact being evaluated [24]. The final sample of 32 students meets the minimum recommended for stable UEQ mean estimates ($n \geq 20-30$) [24]. We acknowledge that 32 students from a single school limits generalisability; findings should be interpreted as applying to comparable TVET contexts rather than Indonesian vocational education broadly [4].

2) Data collection procedure

Each participant completed at least one OOP assessment session in OOPedia that included both the fill-in-the-blank and the radio-button features and then completed the UEQ questionnaire in a single classroom session. UEQ responses were collected on the standard 26-item bipolar scale in the order specified by the official UEQ procedure. At the end of the session, participants were also invited to write a brief open-ended comment in response to the prompt: “What, if anything, did you find unclear or confusing about the assessment features?” Responses were voluntary and unstructured. Approximate total completion time was 12–15 minutes per respondent.

3) Instrument

The UEQ contains 26 bipolar adjective pairs measuring six dimensions: Attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty [12], [24]. Each item is scored on a 7-point Likert scale transformed into a bipolar scale ranging from -3 (most negative) through 0 (neutral) to +3 (most positive), following the official UEQ Data Analysis Tool procedure [24].

The UEQ's validity has been established through its original construction and psychometric evaluation by [12]. It has been confirmed through subsequent cross-cultural applications across multiple languages and contexts [19], [24], [25]. The instrument's factorial structure, separating pragmatic quality (perspicuity, efficiency, dependability) from hedonic quality (stimulation, novelty) with Attractiveness as an overarching valence dimension, has been replicated in independent validation studies [12], [24]. For this study, internal consistency (reliability) for each of the six dimensions was assessed using Cronbach's α , computed from the four items composing each

dimension. A threshold of $\alpha \geq 0.70$ was adopted as the criterion for acceptable reliability [26], while α values between 0.60 and 0.70 were considered acceptable in exploratory research contexts [27]. Dimensions falling below 0.60 are reported with explicit caution in the Results section.

4) Data Analysis

For each UEQ dimension, we computed the mean and variance across respondents, then applied a one-sample t-test against the neutral midpoint ($\mu_0 = 0$) to test whether the dimension mean differs significantly from neutral at $\alpha = 0.05$ ($df = 31$). For each dimension, we also report the 95% confidence interval (CI) of the mean. Internal consistency for each dimension was assessed using Cronbach's α . The resulting dimension means were then compared against the UEQ benchmark intervals reported by [14], which classify mean scores into excellent, good, above average, below average, and bad bands. We note that the inferential statistics were computed from the dimension-level summary statistics; individual item-level responses are not reported in this paper. Open-ended comments were analyzed thematically: responses were read independently by two authors, grouped into recurring themes, and used to triangulate the quantitative UEQ findings, particularly the perspicuity variance.

IV. RESULT

A. Reliability

Five of the six UEQ dimensions showed acceptable internal consistency (Table 1). Cronbach's α ranged from 0.65 (Dependability) to 0.85 (Attractiveness), within the acceptable range for exploratory research [28]. The exception was Novelty ($\alpha = 0.52$), which falls below both the conventional threshold of 0.70 [28] and the exploratory minimum of 0.60 [26]. Novelty scores in subsequent sections should therefore be treated as indicative rather than precise, and are discussed further in Section V.

Table 1.
Reliability of UEQ Dimensions (Cronbach's α , $n = 32$)

Dimension	α	Interpretation
Attractiveness	0.85	Acceptable
Perspicuity	0.69	Acceptable (borderline)
Efficiency	0.76	Acceptable
Dependability	0.65	Acceptable (borderline)
Stimulation	0.72	Acceptable
Novelty	0.52	Below threshold

B. Descriptive Results

All six UEQ dimensions produced positive means, indicating that students responded favorably to OOPedia's assessment features across all quality dimensions (Table II). The two highest scores, stimulation (2.05) and attractiveness (2.03) that indicate that students found the features genuinely engaging and visually appealing, not merely tolerable. Perspicuity (1.60) showed the widest variance (1.61), nearly twice that of attractiveness (0.73), suggesting that the experience of clarity was unevenly distributed across students,

a pattern explored further through open-ended responses in Section IV.F.

C. Inferential Results (One-Sample *t*-tests vs. Neutral Midpoint)

All six dimensions scored significantly above the UEQ neutral midpoint (all $p < 0.001$, Table III), confirming that the positive responses observed in Section IV.B were not attributable to chance in this sample. Statistical significance here is a floor, not a ceiling; the more meaningful question is where each dimension sits within the UEQ benchmark, addressed in Section IV.E.

Table 2.
UEQ Dimension Means and Variances (n = 32)

Dimension	Mean	Variance
Attractiveness	2.03	0.73
Perspicuity	1.60	1.61
Efficiency	1.60	1.07
Dependability	1.64	0.79
Stimulation	2.05	1.07
Novelty	1.33	1.20

Table 3.
One-Sample *t*-tests Against the Neutral Midpoint (df = 31)

Dimension	Mean	SD	SE	t	p	95% CI
Attractiveness	2.03	0.854	0.151	13.454	< 0.001	[1.724, 2.340]
Perspicuity	1.60	1.269	0.224	7.155	< 0.001	[1.148, 2.062]
Efficiency	1.60	1.034	0.183	8.777	< 0.001	[1.232, 1.978]
Dependability	1.64	0.889	0.157	10.419	< 0.001	[1.317, 1.957]
Stimulation	2.05	1.034	0.183	11.200	< 0.001	[1.675, 2.421]
Novelty	1.33	1.095	0.194	6.873	< 0.001	[0.936, 1.726]

D. Pragmatic and Hedonic Quality

Hedonic quality (1.69) marginally exceeded pragmatic quality (1.62), with Attractiveness at 2.03 (Table 4). This pattern is noteworthy: it suggests that students found OOPedia's assessment features motivating and engaging beyond their functional usability, an unusual profile for a tool evaluated primarily on task-completion grounds. Most LMS evaluations show pragmatic quality dominating hedonic quality [23]; the inversion here points to the active-construction format of the fill-in-the-blank feature as a likely driver of hedonic engagement. The internal spread within each quality group also differs sharply. The three pragmatic dimensions cluster tightly (Perspicuity 1.60, Efficiency 1.60, Dependability 1.64; range = 0.04), suggesting that OOPedia's basic usability was perceived consistently across the sample regardless of individual differences. The two hedonic dimensions, by contrast, span a much wider range (Stimulation 2.05, Novelty 1.33; range = 0.72). This asymmetry indicates that while pragmatic quality was uniformly adequate, hedonic quality was carried disproportionately by stimulation rather than by the platform genuinely feeling new, a distinction that matters for design: features can be made more engaging (stimulation) more reliably than they can be made more novel to a population already accustomed to similar interactive quiz interfaces.

Table 4.
Pragmatic and Hedonic Quality

Quality Grouping	Mean
Attractiveness	2.03
Pragmatic Quality	1.62
Hedonic Quality	1.69

E. Benchmark Comparison

Against the UEQ global benchmark [14], OOPedia's assessment features performed competitively across all six dimensions (Table 5, Fig. 5). Stimulation (2.05) and attractiveness (2.03) reached the Excellent band, the top 10% of products in the benchmark dataset, indicating that the combined assessment design motivated and appealed to students at a level comparable to best-in-class consumer products. Dependability (1.64), Efficiency (1.60), and Perspicuity (1.60) reached Good. Novelty (1.33) reached 'Above Average', the only dimension that did not reach 'Good' or higher that consistent with the scaffolding gap identified in the open-ended responses. No dimension fell below the 'Above Average' threshold, meaning OOPedia's combined assessment design meets the benchmark across the full UEQ quality spectrum. In distributional terms, two of six dimensions (33%) reached the Excellent band, three (50%) reached Good, and one (17%) reached Above Average, with none falling into Below Average or Bad. Because the Excellent band represents the top decile of the global UEQ benchmark sample [14], the fact that a third of OOPedia's dimensions placed there is a comparatively strong result for an academic prototype evaluated against a benchmark built primarily from commercial products. No dimension required flagging for poor performance, which narrows the practical design agenda to two targeted concerns rather than a broad usability overhaul: closing the perspicuity variance gap (Section IV.F) and strengthening the novelty construct's reliability (Section V.B).

Table 5.
Benchmark Classification of UEQ Dimensions

Dimension	Mean	Benchmark band
Attractiveness	2.03	Excellent
Perspicuity	1.60	Good
Efficiency	1.60	Good
Dependability	1.64	Good
Stimulation	2.05	Excellent
Novelty	1.33	Above Average

F. Open-Ended Comments

Five of 32 participants provided written comments; all five addressed the fill-in-the-blank feature, none the radio-button. Two themes emerged consistently. The first theme, present in four of five responses, concerned the absence of sufficient contextual cues within the items. Representative responses included: "Some blanks had too few hints, so I was confused about what term to write," and "Sometimes I did not know whether the answer should be code, a keyword, or a short explanation." These comments point to ambiguity about the expected form of the answer, not only the content, but also a scaffolding gap that the current item design does not address.

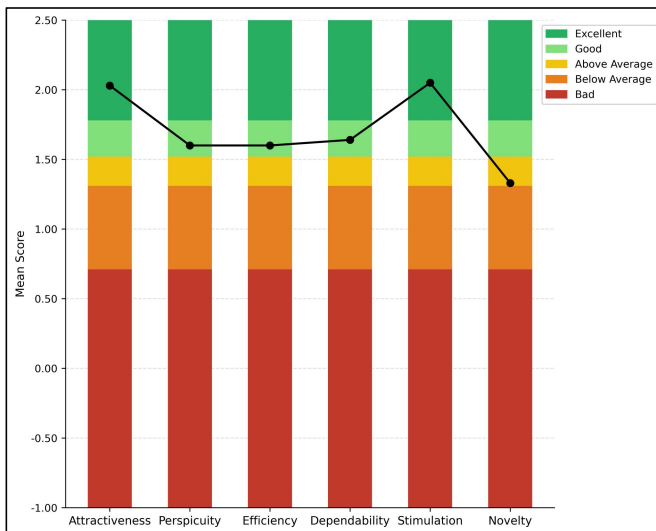


Fig. 5. UEQ benchmark profile of OOPedia's assessment features compared against the global dataset of Schrepp et al. [14], showing dimension means plotted against Excellent, Good, Above Average, Below Average, and Bad benchmark intervals

The second theme, present in three of five responses, concerned prior OOP knowledge as a moderator of clarity. Representative responses included: "The feature is clear enough if you already know the OOP concepts, but I was still confused about new terminology," and "Inheritance material is still hard to understand, so I was often unsure about my fill-in-the-blank answers." One student captured both themes simultaneously: "The interface is easy to use, but some questions were unclear in what they were asking."

These responses are consistent with the Perspicuity variance (1.61) observed in the quantitative data and provide direct evidence that at least a subgroup of students experienced clarity problems attributable to insufficient scaffolding and uneven OOP vocabulary, rather than to the interface design itself.

V. DISCUSSION

A. Interpretation of the Six UEQ Dimensions

The UEQ profile has a clear shape: hedonic quality led pragmatic quality, and the two highest-scoring dimensions (stimulation and attractiveness) both reached the "excellent" band. This pattern is not what typical LMS evaluations produce. Reference [23] found pragmatic quality to dominate hedonic quality across Jordanian Moodle deployments; the inversion here points to something specific about the combined assessment design rather than to educational platforms in general.

The stimulation result is theoretically coherent. Construction-based formats, where learners retrieve and produce rather than select, are associated with greater cognitive engagement than recognition-only tasks [6]. A student completing a fill-in-the-blank prompt must activate and apply OOP knowledge, not merely recognize it among distractors.

That this format produces an excellent stimulation score is consistent with the cognitive engagement literature, though establishing causality would require a controlled comparison with a recognition-only group.

Perspicuity tells a different story. At 1.60 (Good), perspicuity is the only dimension that did not reach the Excellent band, and its variance of 1.61 is the highest among all dimensions, roughly double that of dependability (0.79) and attractiveness (0.73). High variance means the population split: some students found the features clear, others did not. The open-ended responses in Section IV.F resolve the split. Students who struggled pointed to two problems: items that gave no indication of expected answer format (code vs. keyword vs. phrase) and items that probed OOP concepts they had not yet consolidated. Students who found the features clear noted that clarity depended on already knowing the concept. This issue is not an interface problem (the attractiveness and efficiency scores confirm that the interface itself was well received). It is a scaffolding problem in the item design.

Cognitive load theory offers a mechanism [6], that construction-based items transfer processing effort from evaluating distractors to retrieving the target construct. For students with consolidated OOP vocabulary, this is a productive challenge. For students still building that vocabulary, the same demand becomes a barrier. The result is the bimodal perspicuity pattern: not one population with moderate clarity, but two groups with opposite experiences of the same items.

B. The Low Novelty Reliability

The novelty dimension requires careful treatment. The mean (1.33, above average) and the reliability ($\alpha = 0.52$) address different questions and should not be conflated. The positive mean is real: students did perceive the features as novel relative to the neutral midpoint, and the t-test confirms that the result is not sampling noise. The low α is a separate issue: the four novelty items did not co-vary consistently within this sample, which means the composite score cannot be treated as a well-defined single construct.

The most plausible explanation is prior familiarity with the platform. Students in Indonesian secondary schools routinely use Google Classroom, Quizizz, and Kahoot, all of which involve digital quiz interaction. From that baseline, OOPedia's radio-button feature may not be genuinely novel, regardless of its OOP-specific content. Meanwhile, [14] documented this pattern in the UEQ benchmark data: Novelty scores compress when users have prior exposure to functionally similar interfaces. Also, [21] and [22] both observed the same compression in Indonesian educational contexts.

The implication is methodological rather than about OOPedia's design. The four-item UEQ Novelty scale uses generic innovation vocabulary (creative, inventive, and leading-edge) that maps poorly onto a subject-specific educational platform. Future evaluations of OOP learning systems should consider supplementing or replacing these items with domain-specific novelty probes that ask what, if

anything, students found new about the assessment interaction rather than about the platform as a product.

C. Comparison with Related Studies

Positioning OOPedia against comparable UEQ evaluations reveals a consistent pattern and one notable departure. The consistent pattern: The consistent pattern: Novelty compression among Indonesian student samples, observed here and replicated in [21], [22], and implicitly in [19]. Platform familiarity appears to be a structural feature of this population that suppresses the novelty dimension across all systems evaluated.

The notable departure is in the hedonic-pragmatic balance. Prior study, [23] conducted a Moodle evaluation that found pragmatic quality clearly dominating hedonic quality, the expected pattern for a content-delivery and task-management platform. OOPedia inverts this pattern. The most plausible explanation is the nature of the interaction demanded: navigating a Moodle course page is largely passive retrieval of information, while completing a fill-in-the-blank code item requires active knowledge construction. The hedonic uplift in OOPedia's profile likely reflects the engagement that active construction generates, not a property of the interface per se but of the cognitive demand embedded in the assessment format.

The Halodoc comparison [20] further sharpens this point. Halodoc reached only the satisfactory benchmark band overall, despite being a mature, commercially optimized mobile application. OOPedia's stimulation and attractiveness both received excellent ratings, which is a higher result than that of a prototype academic system and a production consumer app. This result does not reflect better interface design; Halodoc almost certainly has a more polished UI. It reflects the motivational properties of active-construction assessment tasks, a finding with direct implications for designers of vocational e-learning systems: assessment interaction type, not interface polish, appears to be the primary lever for hedonic UX in this context.

D. Design Implications

Three design directions follow directly from the findings.

1) Fill-in-the-blank scaffolding

The perspicuity variance and the open-ended comments both point to the same gap: students need to know what form their answer should take before they can engage productively with the content. An optional hint layer, a one-line definition of the OOP term being probed, or a label indicating whether the expected answer is a keyword, a code fragment, or a short phrase that would give students with weaker vocabulary a route into the item without removing the retrieval challenge for students who do not need the scaffold. The hint must remain optional; making it automatic would remove the construction demand that likely drives the high stimulation score.

2) Radio-button feedback depth

The instant feedback mechanism already distinguishes OOPedia's radio-button feature from conventional multiple-choice. However, confirming that an answer was wrong is less useful than explaining which misconception the distractor was designed to surface. One or two sentences

of concept-specific explanation tied to the chosen distractor would close the formative loop more completely [6].

3) Novelty by differentiation, not decoration

Rather than cosmetic feature additions aimed at a higher UEQ novelty score, the more productive direction is to add affordances absent from generic quiz platforms: dynamic class-object diagrams that update as the student fills a blank, or adaptive difficulty selection based on recent performance. Others worked on immersive OOP environments [17], suggesting that domain-specific interactivity, rather than novelty for its own sake, shifts users' perceptions of innovation.

E. Limitations

Several limitations should be kept in mind when interpreting these findings. The sample of 32 students from a single vocational school is small and non-random, so the results apply most directly to comparable SMK contexts with similar student profiles and similar prior exposure to digital quiz platforms; they should not be read as representative of Indonesian vocational education as a whole. The qualitative data were also limited in depth: the five open-ended responses (about 16% of participants) were enough to triangulate the variance seen in Perspicuity but not to characterize the full distribution of student experiences. A dedicated think-aloud protocol with a purposive sub-sample of eight to ten students would support considerably stronger claims about the mechanisms behind the perspicuity split.

Measurement reliability is a further concern: the novelty dimension's internal consistency was low (Cronbach's $\alpha = 0.52$, below both conventional and exploratory thresholds), so its mean (1.33, above average) should be read as a directional signal rather than a reliable point estimate. Finally, the study measured user experience rather than learning outcomes, and a positive UX profile is a necessary but not sufficient condition for learning efficacy. Establishing whether OOPedia's combined assessment design improves OOP learning relative to conventional assessment would require a controlled study with pretest–posttest measures and a matched comparison group, which is the recommended next step.

VI. CONCLUSION

OOPedia's combined fill-in-the-blank and radio-button assessment features produced a positive user experience across all six UEQ dimensions, with stimulation and attractiveness reaching the excellent benchmark band and the remaining four reaching 'Good' or 'Above Average'. The most substantively intriguing finding is not the absolute scores but the pattern: hedonic quality outpaced pragmatic quality, suggesting students were genuinely engaged by the active-construction format rather than merely finding it functional. For an academic prototype evaluated against a global consumer-product benchmark, this finding is a stronger result than the literature on LMS user experience would predict.

Two findings qualify the picture and point to concrete next steps. First, perspicuity showed the widest variance of any dimension, not because the interface was unclear, but because the fill-in-the-blank items lacked scaffolding cues about expected answer format and assumed OOP vocabulary that

some students had not yet consolidated. This condition is a design gap with a known fix: optional on-demand hints and explicit answer-format labels. Second, the novelty scale showed limited internal consistency, meaning its score should be read as a directional signal rather than a reliable estimate. The scale itself likely needs domain-specific revision before it can measure novelty reliably in TVET contexts.

This study evaluated user experience rather than learning outcomes. The UX baseline established here is a necessary precursor, not a substitute, for evidence of learning efficacy. The recommended next step is a controlled experiment comparing OOPedia's combined assessment design against a recognition-only condition, using pretest-posttest OOP learning measures, a matched control group, and a sample drawn from multiple vocational schools across different regions. That design would test whether the engagement advantage reflected in the hedonic UX profile translates into measurable learning gains, which is the question this study was never positioned to answer, but has now laid the groundwork to ask.

Beyond this controlled comparison, two further directions follow from the present results. The design improvements identified here, namely on-demand hints and explicit answer-format labels for the fill-in-the-blank items, should be implemented and re-evaluated to confirm whether they reduce the variance observed in perspicuity. The novelty scale should also be revised and re-validated for TVET contexts, as its low internal consistency in this sample limits the reliability of novelty measurement in this population. Together with the controlled learning-outcome study, these steps would extend the UX baseline established here toward a fuller account of how combined interactive assessment supports vocational OOP learning.

ACKNOWLEDGMENT

The authors thank the Department of Information Technology, the Politeknik Negeri Malang, for their support during the implementation and evaluation of OOPedia, as well as the participating students and teachers at the vocational high school where the evaluation was conducted.

REFERENCES

- [1] E. du Plooy, D. Casteleijn, and D. Franzsen, "Personalized adaptive learning in higher education: A scoping review of key characteristics and impact on academic performance and engagement," *Heliyon*, vol. 10, no. 21, Art. no. e39630, Nov. 2024, doi: 10.1016/j.heliyon.2024.e39630.
- [2] M. F. Contrino, M. Reyes-Millán, P. Vázquez-Villegas, and J. Membrillo-Hernández, "Using an adaptive learning tool to improve student performance and satisfaction in online and face-to-face education for a more personalized approach," *Smart Learning Environments*, vol. 11, Art. no. 6, Feb. 2024, doi: 10.1186/s40561-024-00292-y.
- [3] M. A. Almaiah, A. Al-Khasawneh, and A. Althunibat, "Exploring the critical challenges and factors influencing the E-learning system usage during COVID-19 pandemic," *Educ. Inf. Technol. (Dordr.)*, vol. 25, no. 6, pp. 5261–5280, Nov. 2020, doi: 10.1007/s10639-020-10219-y.
- [4] A. Habibi, S. Sofyan, and A. Mukminin, "Factors affecting digital technology access in vocational education," *Sci. Rep.*, vol. 13, Art. no. 5682, Apr. 2023, doi: 10.1038/s41598-023-32755-6.
- [5] X. Tan, X. Lin, and R. Zhuang, "Development and validation of a secondary vocational school students' digital learning competence scale," *Smart Learning Environments*, vol. 11, Art. no. 37, Aug. 2024, doi: 10.1186/s40561-024-00325-6.
- [6] E. V. van Wijk *et al.*, "Use of very short answer questions compared to multiple choice questions in undergraduate medical students: An external validation study," *PLoS One*, vol. 18, no. 7, Art. no. e0288558, Jul. 2023, doi: 10.1371/journal.pone.0288558.
- [7] J. Bennedsen and M. E. Caspersen, "Failure rates in introductory programming," *ACM Inroads*, vol. 10, no. 2, pp. 30–36, Apr. 2019, doi: 10.1145/3324888.
- [8] J. Bennedsen and M. E. Caspersen, "Failure rates in introductory programming," *ACM SIGCSE Bulletin*, vol. 39, no. 2, pp. 32–36, Jun. 2007, doi: 10.1145/1272848.1272879.
- [9] N. Liberman, C. Beerli, and Y. B.-D. Kolikant, "Difficulties in Learning Inheritance and Polymorphism," *ACM Transactions on Computing Education*, vol. 11, no. 1, pp. 1–23, Feb. 2011, doi: 10.1145/1921607.1921611.
- [10] D. U. Bolliger and F. Martin, "Factors underlying the perceived importance of online student engagement strategies," *Journal of Applied Research in Higher Education*, vol. 13, no. 2, pp. 404–419, May 2021, doi: 10.1108/JARHE-02-2020-0045.
- [11] A. Tsihouridis, S. Xinogalos, and A. Ampatzoglou, "Educational Programming Environments for Enhancing Conceptual Design in the Object-Oriented Paradigm: A Systematic Mapping Study," *Journal of Educational Computing Research*, vol. 62, no. 1, pp. 70–102, Mar. 2024, doi: 10.1177/07356331231203251.
- [12] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *Lecture notes in computer science*, 2008, pp. 63–76. doi: 10.1007/978-3-540-89350-9_6.
- [13] M. Schrepp *et al.*, "On the importance of UX quality aspects for different product categories," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 2, pp. 232–246, Jun. 2023, doi: 10.9781/ijimai.2023.03.001.
- [14] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Construction of a benchmark for the user experience questionnaire (UEQ)," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 4, pp. 40–44, Jan. 2017, doi: 10.9781/ijimai.2017.445.
- [15] L. E. Gutiérrez, C. A. Guerrero, and H. A. López-Ospina, "Ranking of problems and solutions in the teaching and learning of object-oriented programming," *Educ. Inf. Technol. (Dordr.)*, vol. 27, no. 5, pp. 7205–7239, Jun. 2022, doi: 10.1007/s10639-022-10929-5.
- [16] C. Watson and F. W. B. Li, "Failure rates in introductory programming revisited," in *Proceedings of the 2014 conference on Innovation & technology in computer science education - ITiCSE '14*, New York, New York, USA: ACM Press, 2014, pp. 39–44, doi: 10.1145/2591708.2591749.
- [17] K. Sunday, S. Y. Wong, B. O. Samson, and I. T. Sanusi, "Investigating the effect of imikode virtual reality game in enhancing object oriented programming concepts among university students in Nigeria," *Educ. Inf. Technol. (Dordr.)*, vol. 27, no. 5, pp. 6819–6845, Jun. 2022, doi: 10.1007/s10639-022-10886-z.
- [18] H. Keuning, J. Jeurings, and B. Heeren, "A Systematic Mapping Study of Code Quality in Education," in *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*, New York, NY, USA: ACM, Jun. 2023, pp. 5–11, doi: 10.1145/3587102.3588777.
- [19] H. B. Santoso, M. Schrepp, L. M. Hasani, R. Fitriansyah, and A. Setyanto, "The use of user experience questionnaire plus (UEQ+) for cross-cultural UX research: evaluating zoom and learn Quran Tajwid as online learning tools," *Heliyon*, vol. 8, no. 11, Art. no. e11748, Nov. 2022, doi: 10.1016/j.heliyon.2022.e11748.
- [20] M. A. Kushendriawan, H. B. Santoso, P. O. H. Putra, and M. Schrepp, "Evaluating user experience of a mobile health application 'Halodoc' using user experience questionnaire and usability testing," *Jurnal Sistem Informasi*, vol. 17, no. 1, pp. 58–71, Apr. 2021, doi: 10.21609/jsi.v17i1.1063.
- [21] M. A. Maricar, D. Pramana, and D. R. Putri, "Evaluasi penggunaan SLiMS pada E-Library dengan menggunakan user experience question

- (EUQ),” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 2, pp. 319–328, Mar. 2021, doi: 10.25126/jtiik.2021824443.
- [22] A. Pratama, A. Farqi, and E. P. Mandyartha, “Evaluation of user experience in integrated learning information systems using user experience questionnaire (UEQ),” *Journal of Information Systems and Informatics*, vol. 4, no. 4, pp. 1019–1029, Nov. 2022, doi: 10.51519/journalisi.v4i4.394.
- [23] A. M. Saleh, H. Y. Abuaddous, I. S. Alansari, and O. Enaizan, “The evaluation of user experience on learning management systems using UEQ,” *International Journal of Emerging Technologies in Learning (iJET)*, vol. 17, no. 07, pp. 145–162, Apr. 2022, doi: 10.3991/ijet.v17i07.29525.
- [24] M. Schrepp, “User Experience Questionnaire Handbook,” Sep. 2023. Accessed: Jun. 16, 2025. [Online]. Available: <https://www.ueq-online.org/Material/Handbook.pdf>
- [25] M. Schrepp, J. Kollmorgen, and J. Thomaschewski, “A comparison of SUS, UMUX-LITE, and UEQ-S,” *Journal of User Experience*, vol. 18, no. 2, pp. 86–104, 2023.
- [26] G. W. Cheung, H. D. Cooper-Thomas, R. S. Lau, and L. C. Wang, “Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations,” *Asia Pacific Journal of Management*, vol. 41, no. 2, pp. 745–783, Jun. 2024, doi: 10.1007/s10490-023-09871-y.
- [27] J. F. J. Hair, G. T. M. Hult, C. M. Ringle, and M. Sarstedt, *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 3rd ed. SAGE Publications, 2021.
- [28] K. S. Taber, “The use of cronbach’s alpha when developing and reporting research instruments in science education,” *Res. Sci. Educ.*, vol. 48, no. 6, pp. 1273–1296, Dec. 2018, doi: 10.1007/s11165-016-9602-2.