Applied Computer Science 2025, vol. 21, no. 3, pp. 99–116
https://doi.org/10.35784/acs_7625

Submitted: 2025-04-23 | Revised: 2025-06-09 | Accepted: 2025-06-24

Keywords: ERP systems, user experience, UEQ, production engineering, UX research

Anna HAMERA [1]
[1] University of Bielsko-Biala, Poland, hamera.annamaria@gmail.com
[*] Corresponding author: hamera.annamaria@gmail.com

Transforming ERP interfaces in production environments: An empirical evaluation using the User Experience Questionnaire

Abstract

This study aims to evaluate the user experience (UX) of Enterprise Resource Planning (ERP) system modules in Polish enterprises' information and communication technology-based production environments. The research plan includes quantitative research as part of a doctoral thesis, which will be complemented by qualitative methods such as task-based usability tests, heuristic analysis and in-depth interviews with users. A descriptive research design was employed using an online survey incorporating the User Experience Questionnaire (UEQ) to gather quantitative data. The survey was distributed to a diverse group of respondents, including students, alumni, and production practitioners, to capture perceptions of three screens of the ERP system (General Operations Registration, Personalised Operations Registration, and Employee Panel Operations Registration), which enable the registration of operations. Raw responses from a seven-point Likert scale were transformed into a continuous scale and statistical analyses were conducted to compute descriptive metrics and confidence intervals. The findings indicate that the pragmatic dimensions — namely, perspicuity, efficiency, and dependability — received favourable evaluations, demonstrating robust usability and clear functionality. In contrast, the hedonic dimensions, particularly stimulation and novelty, were rated as neutral to negative. This suggests that, although the ERP modules effectively support routine tasks, they lack innovative appeal and engaging design. Benchmark comparisons revealed that the interfaces generally fell within the lower quartile, highlighting the need for targeted UI/UX refinements to enhance visual attractiveness and user motivation. In conclusion, the study highlights the importance of balancing functional performance with improved aesthetic and hedonic attributes to optimise the effectiveness of ERP systems in production engineering settings.

1. INTRODUCTION

Enterprise Resource Planning (ERP) systems have become the cornerstone of modern business operations, particularly in production environments. As integrated platforms, they streamline critical processes, including finance, logistics, production scheduling and human resources, thereby enabling real-time decision-making and operational control. In today's digital era, ERP systems facilitate the 'paperless factory' paradigm (Mleczko, 2016; Matuszek & Kurczyk, 2017) and support efficient production management by integrating complementary systems, such as Manufacturing Execution Systems (MES) and Advanced Planning and Scheduling (APS) (Knosala, 2023). A critical factor in the success of an ERP system is the quality of its user experience (UX). Rigorous UX evaluation is essential because it directly affects training times, error rates and overall operational productivity in production environments (Norman & Nielsen, 1998; Winter, 2025).

While numerous studies (CK & Setiawan, 2021; Munandar & Santoso, 2025) have used tools such as the User Experience Questionnaire (UEQ) to evaluate ERP systems, the majority of these studies have been conducted in academic or governmental contexts. There is a significant gap in the literature concerning the evaluation of ERP interfaces in ICT-driven production environments, where even minor usability enhancements can lead to substantial efficiency improvements in real-world manufacturing processes. A multidimensional analysis of ERP systems — from technical performance to overall user satisfaction — is thus imperative in order to capture the complexity inherent in these industrial applications. This study aims to address this gap by applying the UEQ framework to ERP modules deployed in Polish production enterprises. Specifically, the study evaluates three distinct user interface views: general operations registration, personalised operations registration, and an employee panel registration view. This approach is intended to

comprehensively capture the various aspects of system performance, usability and overall user satisfaction, thereby generating actionable benchmarks for subsequent interface improvements. This will inform both theoretical models and practical strategies for enhancing ERP systems in modern production settings.

The remainder of the article is organised as follows. Chapter 2 provides a detailed literature review, outlining the evolution and significance of ERP systems, the importance of UX in production engineering, and the UEQ's methodological underpinnings. Chapter 3 describes the research methodology, including study design, instrumentation and data analysis techniques. Chapter 4 presents the results alongside benchmark comparisons and reliability assessments for the three UI views. Chapter 5 discusses the results in relation to existing literature, delineating practical implications for ERP system design and user training. Finally, Chapter 6 concludes the study by summarising the key findings, acknowledging the limitations and proposing directions for future research.

2. USER EXPERIENCE IN ERP SYSTEM CONTEXT

Enterprise Resource Planning systems form the backbone of contemporary production management. They integrate critical processes, including finance, logistics, production and human resources, into a unified digital platform that supports decision-making across the entire enterprise (Chomuszko, 2016; Auksztol et al., 2015). In modern production environments, digital transformation, as exemplified by the 'paperless factory' concept, has become indispensable. Integrating ERP with complementary systems, such as Manufacturing Execution Systems (MES) and Advanced Planning and Scheduling (APS), enables real-time process monitoring, optimisation of resources, and reduction of errors in settings with high variability (Mleczko, 2016; Matuszek & Kurczyk, 2017; Knosala, 2023). Furthermore, compliance with international standards such as ISO 9241-210 (International Organization for Standardization, 2010) guarantees that ERP interfaces are designed with a human-centred approach, enhancing ergonomic efficiency and mitigating operator fatigue — crucial factors in optimising production lines and reducing waste (Alves et al., 2023; Antonowicz & Zaborowski, 2023; Fokczyński, 2023; Mleczko, 2015; Parys, 2005; Parys, 2010).

2.1. Significance of UX in ERP systems

User experience (UX) is emerging as a vital determinant of an enterprise resource planning (ERP) system's overall efficacy, particularly in production environments where seamless process execution is paramount. The holistic concept of UX encompasses pragmatic quality dimensions, such as efficiency, clarity, and reliability, as well as hedonic dimensions, including stimulation and novelty (Schrepp, 2015; Hassenzahl & Tractinsky, 2006). Superior UX is directly correlated with reduced training times, minimised error rates and increased productivity. In industrial contexts, where system downtime or user misinterpretation can result in substantial operational losses, it is crucial that ERP interfaces facilitate intuitive, human-centred interactions (Nielsen-Norman & Nielsen, 1998; Winter, 2025). Recent research emphasises that integrating human-centred design, a characteristic of Industry 5.0 paradigms, provides a strategic advantage for production enterprises. For example, Hamera (2024) and Longo et al. (2020) emphasise that a shift towards a human-centric approach enhances user satisfaction and bolsters overall operational performance by aligning system functionality with ergonomic and cognitive requirements. These approaches also promote sustainable practices and reduce inefficiencies, which is crucial for small and medium-sized production enterprises (Gross-Gołacka & Ożga, 2023; Nugraha & Fatwanto, 2021).

2.2. User experience evaluation using the user experience questionnaire

Given the growing emphasis on end-user satisfaction as one of the performance metrics in production engineering, evaluating UX in interactive ERP systems is imperative. The User Experience Questionnaire (UEQ) is a robust, standardised, multidimensional instrument that quantifies both the pragmatic and hedonic aspects of interaction. It achieves this through a semantic differential scale covering six domains: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation and Novelty (Schrepp, 2015). Schrepp et al. (2017) further elaborate on the UEQ's methodological rigour, emphasising its ability to produce reliable benchmarks across a wide array of products. In ERP applications, studies by CK and Setiawan (2021), as well as by Munandar and Santoso (2025), have used the UEQ to evaluate complex systems, including the PeopleSoft Oracle Campus Solution and Indonesia's government financial management system. Although

these studies were conducted in academic or governmental contexts, their findings where dimensions such as dependability, stimulation, and novelty achieved 'excellent' ratings, while attractiveness, efficiency, and perspicuity were rated 'good' demonstrate the UEQ's capacity to capture nuanced user perceptions. This robust evaluative mechanism is crucial in production settings, where even marginal improvements in usability can result in substantial operational gains.

2.3. Research gap and contribution

Notably, while previous studies such as those by Ramadan and Setiawan (2023) and Munandar and Santoso (2025) have used the UEQ to evaluate ERP systems, they have primarily focused on academic or governmental settings rather than ICT-centric production environments. While these studies have provided valuable insights into UX dimensions, they have not addressed the specific requirements of production engineering contexts.

In manufacturing enterprises, where even minor enhancements to system interface and process usability can result in substantial operational improvements, the absence of extensive UEQ evaluations in ICT-oriented production settings represents a significant gap in the literature. Modern production environments require ERP systems that integrate key business functions and facilitate real-time decision-making and seamless human–machine interaction. This is particularly important in light of the digital transformation priorities of Industry 5.0, which emphasises human-centred design alongside technical efficiency (Hamera, 2024; Longo et al. 2020).

The present study aims to address this issue by applying the UEQ framework to ERP modules deployed within ICT-driven production settings in Polish enterprises. By extending the application of the UEQ to these environments, the study provides empirical benchmarks that reflect the operational realities of modern production engineering. Furthermore, the investigation will capture the perspectives of current production practitioners and potential future employees, populations that have largely been overlooked in previous studies. This approach will facilitate iterative improvements to ERP interfaces based on robust user feedback and inform the design of systems that align with the ergonomic and cognitive demands of the production domain.

This research will contribute novel empirical data highlighting the importance of integrating human-centred design principles, such as those recommended by ISO 9241-210 (International Organization for Standardization, 2010) and elucidated by Hassenzahl and Tractinsky (2006), into ERP systems. Integrating these principles is essential to enhancing the overall efficacy of systems, reducing operational errors, and ultimately bolstering the competitiveness and sustainability of production enterprises.

3. METHODOLOGY

3.1. Study design

This study adopts a descriptive research design to address an identified research gap in production engineering. As part of a broader doctoral research plan, an original methodology was developed to support the evolutionary redesign of long-established information systems that are burdened by technological and user experience (UX) debt. This proprietary approach integrates quantitative methods, such as the UEQ (described herein) supplemented by the System Usability Scale, NASA-TLX and the Single Ease Question, and synthesises these with qualitative techniques, such as heuristic analysis, in-depth individual interviews and accessibility assessment. While ERP systems and their associated UX have been extensively researched, prior studies have predominantly focused on academic settings or environments that do not directly reflect ICT-driven production operations. Therefore, this study seeks to augment the existing body of knowledge by focusing specifically on ERP modules deployed within production enterprises, particularly those operating within the Polish manufacturing sector. By doing so, the research aims to provide an empirical examination of user interaction with critical ERP modules, thereby expanding understanding of the relationship between UX phenomena and real-world production processes. This approach is based on the idea that even small improvements in system usability can lead to significant increases in operational efficiency and reductions in errors — critical factors in production engineering. This methodological choice is also justified by the need to document and quantitatively assess user perceptions to guide iterative interface design and support decision-making processes that are integral to maintaining competitiveness and sustainability in the modern production landscape. These studies form part of a broader research plan and represent a pilot phase designed to validate

the methodology under specific conditions. The overall plan also includes a subsequent UX research programme involving a cohort of production workers with prior ERP system experience.

3.2. Instrumentation

This study primarily employs the User Experience Questionnaire (UEQ), a robust and extensively validated tool originally developed in Germany in 2005. The UEQ is specifically designed to capture the holistic concept of user experience (UX) by evaluating both the pragmatic and hedonic aspects of interactive systems (Laugwitz et al., 2008). Its popularity and utility in academic and industrial research stem from its capacity for rapid deployment and extensive benchmarking capabilities. In the current study, the means of the measured scales are being compared against a comprehensive benchmark dataset comprising data from 21,175 respondents across 468 evaluation studies. This dataset covers a wide range of products, including business software, websites, web shops and social networks, thereby providing a reliable basis for interpreting the quality of the evaluated ERP module.

- The UEQ consists of 26 items, which are divided into six distinct scales:
- Attractiveness: Capturing the product's overall impression and emotional appeal.
- Perspicuity: Assessing how quickly users can become familiar with and learn to operate the system.
- Efficiency: Measuring the system's ability to support task completion with minimal effort.
- Dependability: Evaluating the extent to which users feel in control during interaction.
- Stimulation: Indicating whether the product is exciting and motivating to use.
- Novelty: Reflects the product's creative and innovative aspects.

Each item is presented as a semantic differential, in which respondents rate the product on a seven-point Likert scale that has been converted into a continuous scale ranging from –3 (most negative) to +3 (most positive).

This method reduces central tendency bias and enables more nuanced measurement of subjective user experiences. The scale distinguishes not only between goal-oriented, pragmatic qualities, such as perspicuity, efficiency and dependability, but also hedonic qualities, namely stimulation and novelty, that are non-goal directed (Schrepp, 2015). The instrument’s reliability and construct validity have been corroborated through multiple usability tests involving 144 participants in controlled environments, as well as an online survey of 722 participants. These studies have demonstrated high internal consistency, as evidenced by Cronbach's alpha values, and robust construct validity across various application domains.

In addition to the UEQ, supplementary qualitative instruments (e.g. semi-structured interviews and system performance metrics) were employed to contextualise the quantitative findings. However, the primary emphasis is placed on the UEQ due to its well-established use and pivotal role in enabling comprehensive benchmarking against an extensive body of prior research. This multifaceted approach is particularly valuable in production engineering, where meticulous evaluation of user interactions with ERP systems can lead to significant enhancements in process optimisation and operational decision-making.

3.3. Procedure

The survey was conducted during academic classes at the University of Bielsko-Biala. Data were collected via an online survey administered through Google Forms. The survey incorporated the UEQ as a core component of a broader questionnaire designed to provide a holistic evaluation of a web-based ERP module. Aimed at production enterprises within the Polish SME sector, the module supports various operational functions, including the registration of production activities. The ERP module under study offers three distinct UI views: (1) A simplified general registration interface with a basic form, shown in Fig. 1.; (2) A personalised operational registration interface featuring additional data retrieval functionalities, shown in Fig. 2.; (3) A contemporary employee dashboard interface incorporating advanced components such as real-time performance charts and statistical displays, shown in Fig. 3.

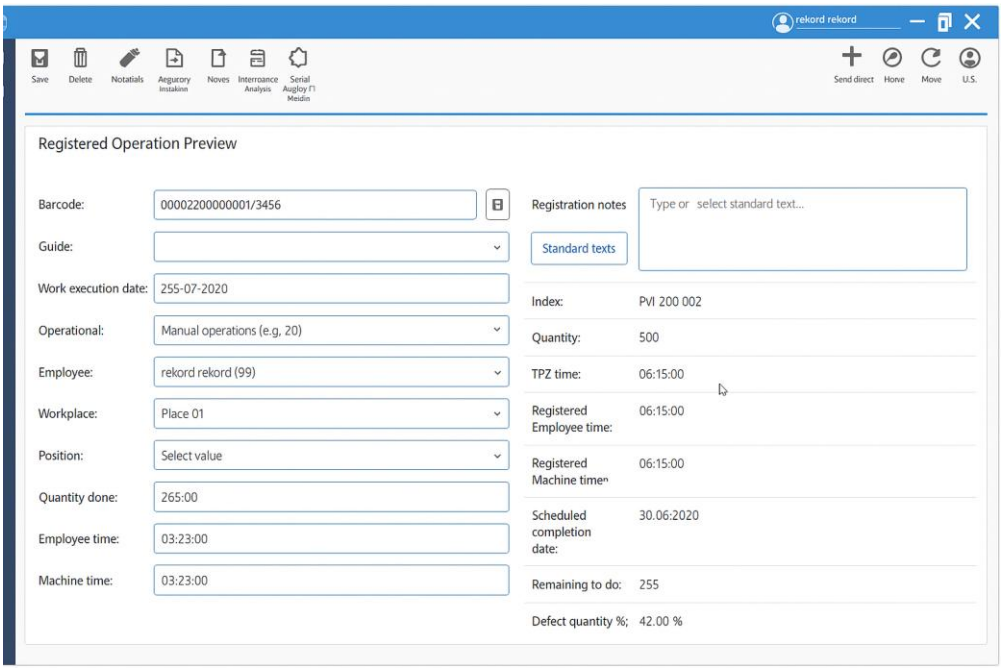


Fig. 1. The general operations registration view

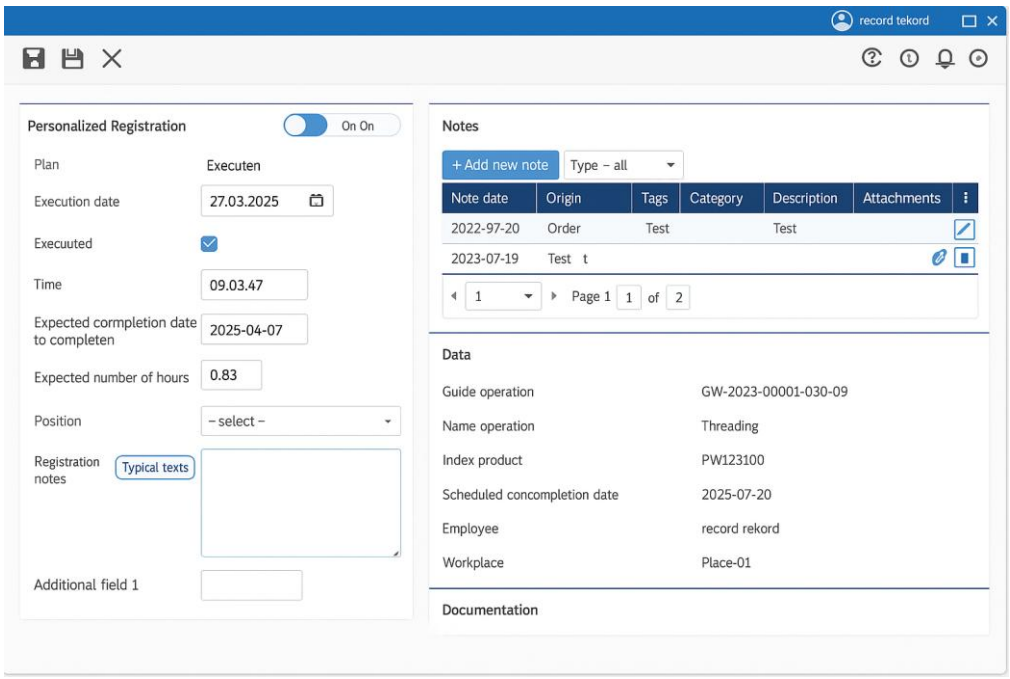


Fig. 2. The personalized operations registration view

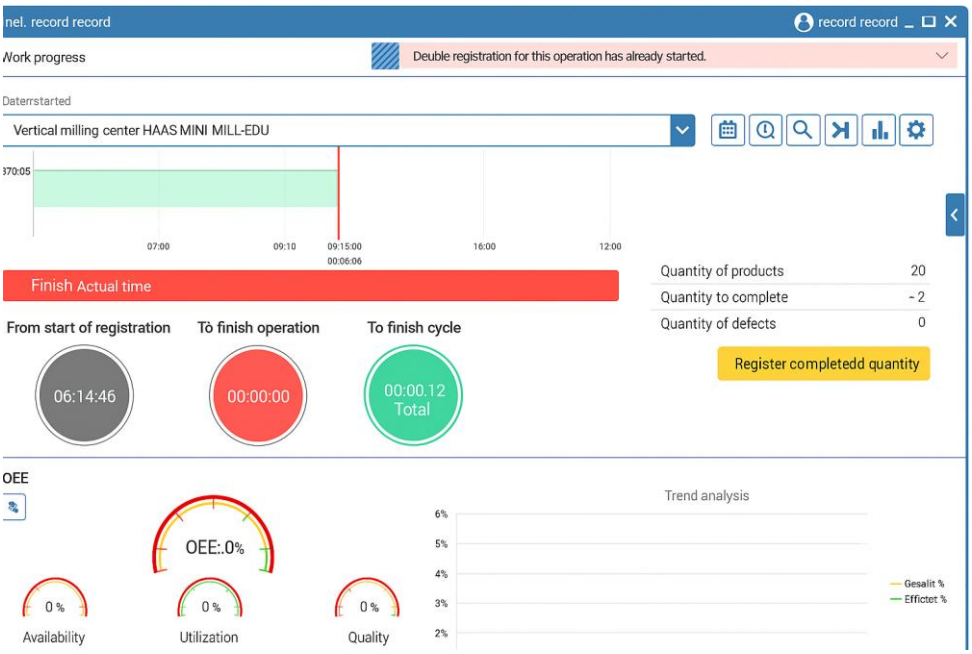


Fig. 3. The employee panel operations registration view

Prior to the full-scale study, a pilot investigation involving 15 respondents was conducted to evaluate the clarity and reliability of the instrument. Feedback from this pilot phase revealed some ambiguities in the official English-to-Polish translation of the UEQ scales. To minimise respondent confusion, therefore, the scales were re-translated and refined by the author of this publication. The final survey was distributed to a diverse cohort of 315 participants comprising students and alumni from the University of Bielsko-Biala. These participants were drawn from undergraduate, master's and doctoral programmes, as well as from non-student populations, and were predominantly from disciplines such as informatics, management and production engineering, and mechanical engineering. This targeted sampling approach enabled data to be collected from participants with varying levels of practical experience with ERP systems, thereby enhancing the study's validity and relevance in production environments.

3.4. Data analysis techniques

The raw scores obtained using the original seven-point Likert scale were imported into the official analytical tool provided by the creators of the UEQ: "UEQ_Data_Analysis_Tool_Version12.xlsx". Within this tool, the raw scores are transformed into a continuous scale ranging from –3 to +3 in accordance with the UEQ guidelines, thereby ensuring compatibility with established benchmark values. The tool then computes key descriptive statistics, including mean scores, standard deviations and 95% confidence intervals, for each dimension measured. These confidence intervals facilitate precise benchmarking by situating the evaluated product relative to reference categories derived from a comprehensive dataset. The tool also calculates reliability metrics such as Cronbach's alpha and Guttman's lambda², which confirm the internal consistency of the individual scales. Combining these techniques, as recommended by Schrepp et al. (2017) allows for a rigorous statistical evaluation of UX. This analytical framework is designed to aggregate user responses, detect trends and identify potential outlier items that may require further examination. It also helps to ensure that each dimension of the UEQ is accurately measured, thereby supporting robust comparative analyses. The actionable insights generated by this process inform iterative design improvements to ERP system interfaces, ultimately contributing to more efficient production processes in SME environments.

3.5. Limitations

Although this study's methodological framework is designed to provide comprehensive insights into the user experience (UX) of ERP systems in production environments, several limitations should be noted. Firstly, reliance on self-reported survey data may introduce subjectivity inherent in user perceptions. Secondly, the timing of data collection, particularly during evening classes, may have affected respondent engagement levels,

potentially attenuating the accuracy of certain responses. Future studies should consider conducting evaluations during periods of peak alertness, such as morning or early afternoon sessions. Lastly, while the current sample size of 315 respondents includes both experienced ERP users and individuals with predominantly academic exposure, further research involving a larger, more diverse cohort (including a higher proportion of production employees) is recommended to enhance the generalisability of the findings.

4. RESULTS

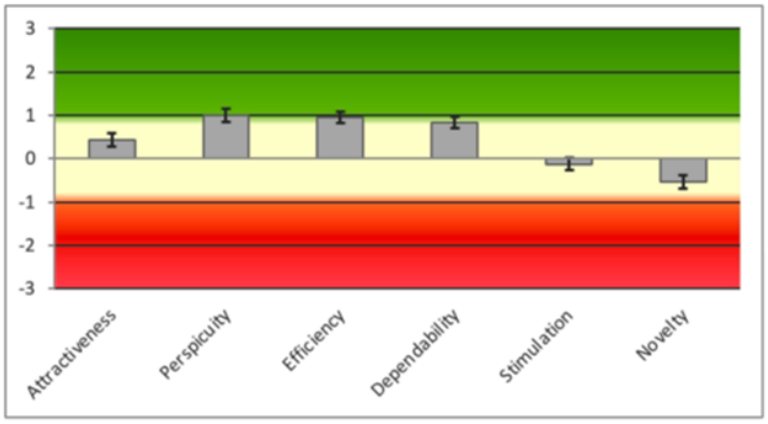
4.1. Six scales results

The six UEQ dimensions (Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation and Novelty) were analysed by interpreting the mean scores for each scale individually, rather than aggregating them into an overall UX metric. Due to the multidimensional structure of the UEQ, it is not methodologically sound to construct a single composite score by averaging all scales, as this fails to preserve the distinct evaluative nuances of each dimension. Instead, the means of each individual scale serve as robust indicators, enabling potential outliers to be identified if an item displays significant deviations from its corresponding scale's overall pattern. Such anomalies may suggest context-specific misinterpretations by respondents.

Values within the interval of –0.8 to +0.8 are generally considered to be indicative of a neutral evaluation. Scores exceeding +0.8 denote a positive appraisal and those below –0.8 reflect a negative perception. While the theoretical scale ranges from –3 (extremely poor) to +3 (extremely good), empirical observations often reveal a narrower range, typically between –2 and +2, due to respondents' tendency to avoid extreme ratings. Consequently, a mean score of +1.5 may appear modest when compared to the full theoretical range, yet it represents a commendable outcome within the context of aggregated respondent data.

4.1.1. General operations registration view

Analysis of the UEQ data (Fig. 5) reveals that certain scale items exhibit distinctly high mean values, particularly in the Efficiency and Perspicuity dimensions. These consistently exceed +1.5, indicating robustly positive assessments of the product's usability and clarity. In contrast, the “usual/leading edge” item registers a mean of approximately –0.8, suggesting that respondents perceive the product as conventional rather than innovative. This divergence suggests that, although the product effectively meets functional and operational requirements, its design may not convey cutting-edge technological innovation. The moderate standard deviations observed across these items imply a reasonable consensus among respondents, although some variability remains that could reflect context-specific interpretations.



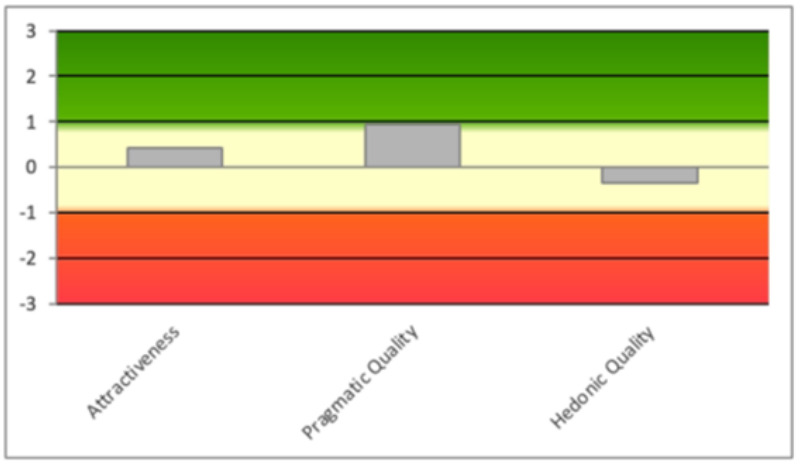
UEQ Scales (Mean and Variance)		
Attractiveness	0,432	1,69
Perspicuity	1,001	1,71
Efficiency	0,949	1,58
Dependability	0,839	1,39
Stimulation	-0,126	1,53
Novelty	-0,539	1,90

Fig. 4. Chart of results of UEQ scales for general operations registration view

Item	Mean	Variance	Std. Dev.	No.	Left	Right	Scale	
1	0,7	2,2	1,5	315	annoying	enjoyable	Attractiveness	
2	1,1	2,4	1,5	315	not understandable	understandable	Perspicuity	
3	-0,5	2,3	1,5	315	creative	dull	Novelty	
4	1,0	2,2	1,5	315	easy to learn	difficult to learn	Perspicuity	
5	0,6	1,9	1,4	315	valuable	inferior	Stimulation	
6	-0,7	2,2	1,5	315	boring	exciting	Stimulation	
7	-0,4	2,3	1,5	315	not interesting	interesting	Stimulation	
8	1,1	2,1	1,5	315	unpredictable	predictable	Dependability	
9	0,6	2,1	1,4	315	fast	slow	Efficiency	
10	-0,6	2,6	1,6	315	inventive	conventional	Novelty	
11	0,7	1,9	1,4	315	obstructive	supportive	Dependability	
12	0,7	2,1	1,4	315	good	bad	Attractiveness	
13	0,9	2,6	1,6	315	complicated	easy	Perspicuity	
14	0,2	2,1	1,4	315	unlikable	pleasing	Attractiveness	
15	-0,8	2,5	1,6	315	usual	leading edge	Novelty	
16	0,4	2,1	1,5	315	unpleasant	pleasant	Attractiveness	
17	0,9	1,8	1,3	315	secure	not secure	Dependability	
18	0,0	2,2	1,5	315	motivating	demotivating	Stimulation	
19	0,7	2,3	1,5	315	meets expectations	does not meet expectations	Dependability	
20	0,8	2,0	1,4	315	inefficient	efficient	Efficiency	
21	1,1	2,3	1,5	315	clear	confusing	Perspicuity	
22	1,1	2,1	1,5	315	impractical	practical	Efficiency	
23	1,3	2,2	1,5	315	organized	cluttered	Efficiency	
24	0,0	2,7	1,6	315	attractive	unattractive	Attractiveness	
25	0,6	2,2	1,5	315	friendly	unfriendly	Attractiveness	
26	-0,3	2,5	1,6	315	conservative	innovative	Novelty	

Fig. 5. Results table of all items for general operations registration view (Schrepp, 2015)

Preliminary analysis of the aggregated UEQ data (Fig. 4) suggests that Attractiveness (mean = 0.432, variance = 1.69) is in a marginally positive range, though it does not exceed the +0.8 threshold, which is often considered to indicate a favourable evaluation. Meanwhile, both Perspicuity (mean = 1.001, variance = 1.71) and Efficiency (mean = 0.949, variance = 1.58) exceed +0.8, suggesting that respondents generally perceived the system as being easy to learn and effective for completing tasks. Dependability (mean = 0.839, variance = 1.39) also hovers above +0.8, implying that users feel reasonably in control of their interactions. In contrast, the hedonic dimensions — Stimulation (mean = -0.126, variance = 1.53) and Novelty (mean = -0.539, variance = 1.90) — score below zero, reflecting neutral-to-unfavourable impressions regarding the product's excitement and innovativeness. Collectively, these results suggest that, although the pragmatic qualities (perspicuity, efficiency and dependability) are viewed favourably, the system's hedonic appeal, particularly its stimulation and novelty, could be enhanced to better capture user interest and motivation. The UEQ scales can be grouped into pragmatic quality (perspicuity, efficiency, dependability) and hedonic quality (stimulation, originality). Pragmatic quality describes task-related quality aspects and hedonic quality describes non-task-related quality aspects.



Pragmatic and Hedonic Quality	
Attractiveness	0,43
Pragmatic Quality	0,93
Hedonic Quality	-0,33

Fig. 6. Chart of results of pragmatic and hedonic quality scales for general operations registration view

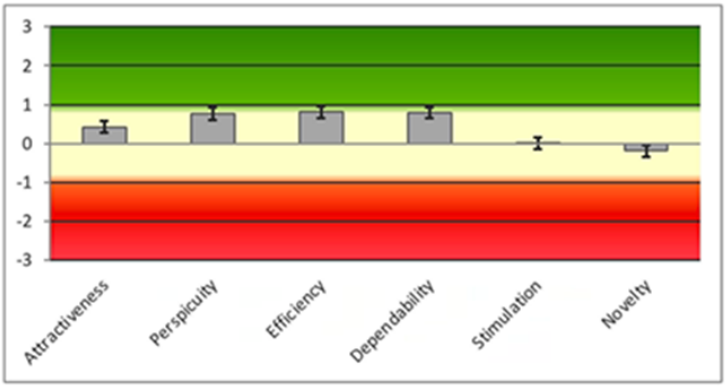
As shown in Fig. 6, Attractiveness yields a mean of +0.43, placing it within the neutral-to-slightly-positive range, though still below the +0.8 threshold required for a distinctly favourable impression. In contrast, pragmatic quality registers at +0.93, indicating a positive evaluation of goal-directed attributes such as perspicuity, efficiency, and dependability. In other words, users generally perceive the system's core functionalities as straightforward, consistent and reliable. Meanwhile, Hedonic Quality stands at –0.33, indicating a neutral but slightly less enthusiastic view of non-task-oriented aspects such as Stimulation and Originality. While this value does not denote a negative impression, it indicates that the system does not capture user interest strongly from an innovative or creative standpoint. Overall, these findings suggest a design that excels in task-focused dimensions, but which could be improved by enhancing its hedonic appeal to improve its overall UX.

4.1.2. Personalized operations registration view

Item	Mean	Variance	Std. Dev.	No.	Left	Right	Scale
1	0,4	2,5	1,6	315	annoying	enjoyable	Attractiveness
2	0,7	2,6	1,6	315	not understandable	understandable	Perspicuity
3	-0,1	2,3	1,5	315	creative	dull	Novelty
4	0,9	2,2	1,5	315	easy to learn	difficult to learn	Perspicuity
5	0,6	1,8	1,4	315	valuable	inferior	Stimulation
6	-0,4	2,4	1,6	315	boring	exciting	Stimulation
7	-0,2	2,5	1,6	315	not interesting	interesting	Stimulation
8	0,9	2,0	1,4	315	unpredictable	predictable	Dependability
9	0,7	2,3	1,5	315	fast	slow	Efficiency
10	-0,2	2,4	1,5	315	inventive	conventional	Novelty
11	0,6	2,0	1,4	315	obstructive	supportive	Dependability
12	0,6	2,2	1,5	315	good	bad	Attractiveness
13	0,7	2,6	1,6	315	complicated	easy	Perspicuity
14	0,3	1,9	1,4	315	unlikable	pleasing	Attractiveness
15	-0,4	2,4	1,6	315	usual	leading edge	Novelty
16	0,4	2,2	1,5	315	unpleasant	pleasant	Attractiveness
17	0,9	1,9	1,4	315	secure	not secure	Dependability
18	0,1	2,0	1,4	315	motivating	demotivating	Stimulation
19	0,8	2,2	1,5	315	meets expectations	does not meet expectations	Dependability
20	0,8	2,1	1,4	315	inefficient	efficient	Efficiency
21	0,8	2,4	1,5	315	clear	confusing	Perspicuity
22	0,9	2,2	1,5	315	impractical	practical	Efficiency
23	0,8	2,7	1,7	315	organized	cluttered	Efficiency
24	0,2	2,5	1,6	315	attractive	unattractive	Attractiveness
25	0,6	2,4	1,5	315	friendly	unfriendly	Attractiveness
26	-0,1	2,7	1,6	315	conservative	innovative	Novelty

Fig. 7. Results table of all items for personalized operations registration view (Schrepp, 2015)

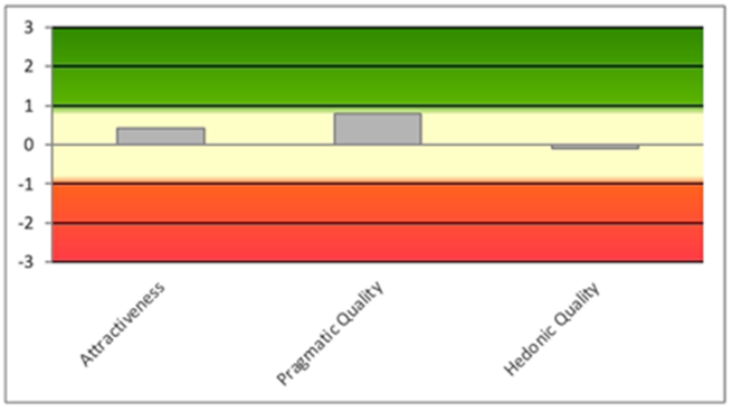
Analysis of the scores (see Fig. 7) indicates that several items achieved distinctly positive mean values above +2.0, for example “not understandable/understandable” and “unpredictable/predictable”, both at 2.5, and “conservative/innovative”, peaking at 2.7. These results suggest that users perceive the interface as clear, predictable and forward-thinking. Conversely, a few items register negatively, such as 'boring/exciting' (-1.1) and 'demotivating/motivating' (-0.9), implying that certain aspects of the product fail to engage or inspire users. Additionally, the borderline mean of –0.8 for 'bad/good' is noteworthy, as it signals an unfavourable impression of the product's overall quality. Nevertheless, the moderate standard deviations observed across most items point to a relatively consistent user consensus, which reinforces the validity of these findings.



UEQ Scales (Mean and Variance)		
Attractiveness	0,422	1,74
Perspicuity	0,767	1,86
Efficiency	0,806	1,84
Dependability	0,785	1,49
Stimulation	0,008	1,64
Novelty	-0,192	1,94

Fig. 8. Chart of results of UEQ scales for personalized operations registration view

Examining the mean scores for each UEQ dimension (Fig. 8) reveals that Attractiveness (0.422) and Perspicuity (0.767) both fall within the slightly positive to borderline range. This suggests that respondents perceive the interface as moderately appealing and relatively easy to learn. Efficiency (0.806) and Dependability (0.785) approach or marginally exceed the +0.8 threshold for a 'positive' interpretation, indicating a favourable assessment of the system's ability to solve tasks and its perceived reliability. Stimulation (0.008) hovers near zero, signifying that users have a largely neutral impression of the interface's motivational or engaging aspects. In contrast, Novelty (-0.192) is in the mildly negative range, suggesting that participants did not perceive the application as particularly innovative or creative. Overall, these results highlight the system's strengths in terms of practicality and consistency (Efficiency and Dependability), while also revealing areas (Stimulation and Novelty) that could benefit from further design refinements to create a more engaging and innovative user experience (UX).



Pragmatic and Hedonic Quality	
Attractiveness	0,42
Pragmatic Quality	0,79
Hedonic Quality	-0,09

Fig. 9. Chart of results of pragmatic and hedonic quality scales for personalized operations registration view

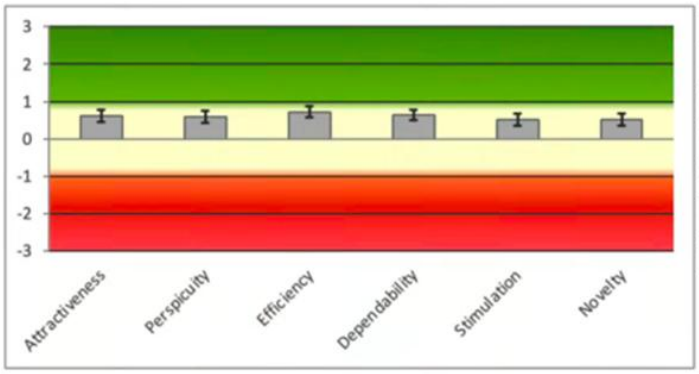
Based on the results shown in Fig. 9, attractiveness yields a mean of +0.42, suggesting an acceptable, though not exceptional, emotional response. In contrast, pragmatic quality stands at +0.79, indicating a favourable evaluation of utilitarian dimensions such as perspicuity, efficiency, and dependability. Meanwhile, Hedonic Quality registers at -0.09, reflecting a neutral stance on non-task aspects such as innovativeness and engagement. These results imply that, while the system competently fulfils its core, goal-oriented functions, it does not captivate users with novel or stimulating features. Consequently, efforts to enhance hedonic elements could further improve the overall user experience (UX).

4.1.3. Employee panel operations registration view

Item	Mean	Variance	Std. Dev.	No.	Left	Right	Scale
1	0,7	3,0	1,7	315	annoying	enjoyable	Attractiveness
2	0,6	2,8	1,7	315	not understandable	understandable	Perspicuity
3	0,6	2,8	1,7	315	creative	dull	Novelty
4	0,6	2,5	1,6	315	easy to learn	difficult to learn	Perspicuity
5	0,7	2,4	1,5	315	valuable	inferior	Stimulation
6	0,4	2,6	1,6	315	boring	exciting	Stimulation
7	0,6	2,6	1,6	315	not interesting	interesting	Stimulation
8	0,5	2,2	1,5	315	unpredictable	predictable	Dependability
9	0,7	2,1	1,5	315	fast	slow	Efficiency
10	0,5	2,7	1,6	315	inventive	conventional	Novelty
11	0,7	2,3	1,5	315	obstructive	supportive	Dependability
12	0,7	2,4	1,6	315	good	bad	Attractiveness
13	0,5	3,0	1,7	315	complicated	easy	Perspicuity
14	0,5	2,3	1,5	315	unlikable	pleasing	Attractiveness
15	0,4	2,6	1,6	315	usual	leading edge	Novelty
16	0,6	2,6	1,6	315	unpleasant	pleasant	Attractiveness
17	0,7	2,1	1,4	315	secure	not secure	Dependability
18	0,4	2,5	1,6	315	motivating	demotivating	Stimulation
19	0,7	2,4	1,5	315	meets expectations	does not meet expectations	Dependability
20	0,7	2,3	1,5	315	inefficient	efficient	Efficiency
21	0,7	2,5	1,6	315	clear	confusing	Perspicuity
22	0,8	2,3	1,5	315	impractical	practical	Efficiency
23	0,6	2,6	1,6	315	organized	cluttered	Efficiency
24	0,6	2,6	1,6	315	attractive	unattractive	Attractiveness
25	0,6	2,7	1,6	315	friendly	unfriendly	Attractiveness
26	0,6	2,7	1,6	315	conservative	innovative	Novelty

Fig. 10. Results table of all items for employee panel operations registration view (Schrepp, 2015)

Examining the UEQ item means (Fig. 10) reveals that several indicators are strongly positive, with the following items exceeding a mean of +2.3: “annoying/enjoyable”, “not understandable/understandable”, “creative/dull”, “boring/exciting”, “complicated/easy”, “impractical/practical”, “unfriendly/friendly” and “conservative/innovative”. These high values imply that respondents find the interface pleasant, intelligible and innovative. In contrast, items such as 'inferior/valuable', 'unattractive/attractive', 'unlikable/pleasing', 'bad/good', and 'demotivating/motivating' have means ranging from –1.0 to –0.8, suggesting shortcomings in perceived value and motivational appeal. The moderate to high standard deviations (mostly 1.5–1.7) reflect variability in user opinions, but do not fundamentally undermine the interpretation of the data as a whole. Overall, the results highlight areas of strength, such as clarity and friendliness, as well as areas for improvement, particularly regarding perceived value and user motivation, providing valuable guidance for targeted interface refinements.

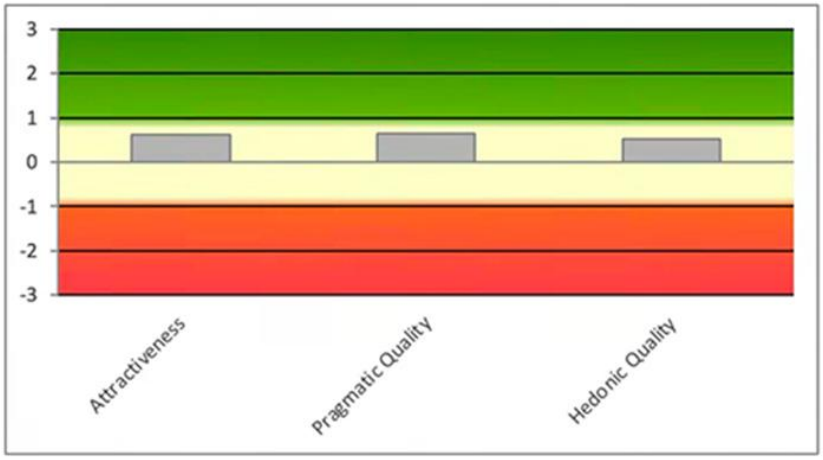


UEQ Scales (Mean and Variance)		
Attractiveness	0,606	2,12
Perspicuity	0,590	2,10
Efficiency	0,717	1,85
Dependability	0,638	1,69
Stimulation	0,520	2,01
Novelty	-0,192	1,94

Fig. 11. Chart of results of UEQ Scales for employee panel registration view

Figure 11 shows the average UEQ ratings for the six core dimensions, indicating that all mean values are between +0.5 and +0.7. Specifically, the highest mean (0.717) and lowest variance (1.85) are attained by Efficiency, suggesting that users found the system’s task execution process relatively straightforward and consistent. Attractiveness and Dependability likewise occupy a moderately positive zone with mean scores of

0.606 and 0.638 respectively, indicating that the interface was perceived as visually acceptable and provided a reasonable sense of control. Perspicuity (0.590) and Stimulation (0.520) both remain just below the +0.6 mark, implying a fair degree of clarity and moderate enjoyment. Finally, Novelty, at 0.525, ranks slightly below the other pragmatic scales, suggesting that users viewed the system's design as functional but not particularly innovative. Nevertheless, none of the scales fell below the neutral threshold of -0.8, reflecting a generally favourable, albeit not highly distinctive, UX.



Pragmatic and Hedonic Quality	
Attractiveness	0,61
Pragmatic Quality	0,65
Hedonic Quality	0,52

Fig. 12. Chart of results of pragmatic and hedonic quality scales for employee panel registration view

In the pragmatic and hedonic quality evaluation (Fig. 12), attractiveness achieves a mean score of +0.61, indicating moderately positive user impressions at an emotional level. Pragmatic Quality rises to +0.85, highlighting the system's ongoing strength in core functionalities and usability (e.g. clarity, efficiency, reliability). Additionally, the Hedonic Quality score of +0.52 suggests a more favourable perception of non-task dimensions, reflecting an improved sense of novelty and stimulation compared to previous measurements. Overall, these data indicate balanced advancement: while essential, goal-directed interactions remain well supported, the system is now also perceived as more engaging and innovative, thus better catering to users' hedonic expectations.

4.2. Confidence intervals for scales

Confidence intervals are a fundamental indicator of the precision of estimated mean values for each scale. In this analysis, 95% confidence intervals were adopted, corresponding to a significance level of $p = 0.05$. This indicates that there is a 95% probability that the true parameter value lies within the specified range. A narrower confidence interval reflects greater precision in the mean estimate and attests to the consistency of respondents' evaluations. The width of the confidence interval is influenced by the sample size and the variability of the data; larger sample sizes and lower dispersion result in a tighter interval. Analysing these intervals enables the stability and reliability of the measured scales to be assessed, which is crucial when comparing product ratings against established benchmark data. Finally, identifying discrepancies in the confidence intervals can highlight potential anomalies, indicating the necessity to scrutinise specific questionnaire items further (Simundic, 2008).

Tab. 1. Confidence intervals per scale for general operations registration view

Confidence intervals (p=0.05) per scale						
Scale	**Mean**	**Std. Dev.**	**N**	**Confidence**	**Confidence interval**	
Attractiveness	0.432	1.298	315	0.143	0.288	0.575
Perspicuity	1.001	1.309	315	0.145	0.856	1.145
Efficiency	0.949	1.255	315	0.139	0.811	1.088
Dependability	0.839	1.181	315	0.130	0.708	0.969
Stimulation	-0.126	1.235	315	0.136	-0.263	0.010
Novelty	-0.539	1.378	315	0.152	-0.691	-0.387

Tab. 2. Confidence intervals per scale for personalized operations registration view

Confidence intervals (p=0.05) per scale						
Scale	**Mean**	**Std. Dev.**	**N**	**Confidence**	**Confidence interval**	
Attractiveness	0.422	1.319	315	0.146	0.277	0.568
Perspicuity	0.767	1.364	315	0.151	0.616	0.917
Efficiency	0.806	1.358	315	0.150	0.656	0.955
Dependability	0.785	1.219	315	0.135	0.650	0.920
Stimulation	0.008	1.280	315	0.141	-0.133	0.149
Novelty	-0.192	1.391	315	0.154	-0.346	-0.038

Tab. 3. Confidence intervals per scale for employee panel operations registration view

Confidence intervals (p=0.05) per scale						
Scale	**Mean**	**Std. Dev.**	**N**	**Confidence**	**Confidence interval**	
Attractiveness	0.606	1.455	315	0.161	0.446	0.767
Perspicuity	0.590	1.450	315	0.160	0.430	0.751
Efficiency	0.717	1.360	315	0.150	0.567	0.868
Dependability	0.638	1.301	315	0.144	0.494	0.782
Stimulation	0.520	1.417	315	0.157	0.363	0.676
Novelty	0.525	1.497	315	0.165	0.359	0.690

Table 1 shows that the General Operations Registration View scores highest on perspicuity (mean = 0.831, 95% CI [0.590, 1.072]), indicating a user-friendly interface. However, efficiency (mean = 0.190, 95% CI [–0.004, 0.384]) suggests only a marginally above-average ease of task completion. 1 shows that the General Operations Registration View has the highest Perspicuity score (mean = 0.831, 95% CI [0.590, 1.072]), indicating a user-friendly interface. In contrast, the Efficiency score (mean = 0.190, 95% CI [–0.004, 0.384]) suggests that the ease of task completion is only marginally above average. The moderate attractiveness score (mean = 0.432) reflects balanced aesthetic appeal. Variations in the width of the confidence intervals indicate differences in the stability of the measurements across scales.

In Tab. 2, the Personalised Operations Registration View demonstrates strong pragmatic dimensions, with dependability and efficiency yielding means of 0.806 and 0.843, respectively. However, the Novelty score is negative (mean = –0.319, 95% CI [–0.570, –0.068]), implying a lack of innovative features. Similar trends are evident in the Employee Panel Operations Registration View (Table 3), with dependability remaining high (mean = 0.624) and novelty lower again (mean = –0.152). Overall, the stable confidence intervals (p = 0.05) suggest that the sample size adequately captures user perceptions.

These findings highlight the importance of balancing functional effectiveness with enhanced hedonic features. While all three views are favourably rated for clarity and reliability, the lower scores for stimulation and novelty highlight opportunities for design improvements that could further engage users and enhance overall system performance.

4.3. Correlations of the items per scale and reliability coefficients

Items that are aggregated into a single scale are generally expected to exhibit high inter-item correlations, indicating that they measure a shared construct coherently (Cronbach, 1951; Schmitt, 1996). In this analysis, the correlations among items allocated to each UEQ scale were examined alongside classical reliability indices, such as Cronbach's alpha and Guttman's lambda². While Cronbach's Alpha is widely cited, its numerical threshold (e.g. 0.7) should be interpreted with caution, particularly for small sample sizes, as these can inflate sampling error (Schmitt, 1996; Bonett, 2002). Deviations from the expected value of Alpha (e.g. below 0.6 or 0.7) could signal potential misinterpretations of specific items or contextual inconsistencies. Furthermore, Guttman's Lambda² offers a more robust lower-bound estimate of true reliability than Alpha does, although minor computational rounding discrepancies may occasionally result in Lambda² being lower than Alpha (Callender & Osburn, 2005). It is important to note that extremely low reliability values in any scale warrant closer inspection of item-to-item correlations, as they may indicate genuine conceptual heterogeneity or contextual biases in item interpretation.

Tab. 4. Reliability indicators for general operations registration view

	Attractiveness	Perspicuity	Efficiency	Dependability	Stimulation	Novelty
Alpha	0.94	0.87	0.88	0.85	0.86	0.89
Conf. Int.	0.92	0.84	0.86	0.82	0.84	0.87
Alpha (5%)	0.95	0.89	0.90	0.87	0.89	0.91
Lambda1	0.78	0.65	0.66	0.64	0.65	0.67
Lambda2	0.93	0.87	0.88	0.85	0.87	0.89

Tab. 5. Reliability indicators for personalized operations registration view

	Attractiveness	Perspicuity	Efficiency	Dependability	Stimulation	Novelty
Alpha	0.94	0.90	0.91	0.88	0.89	0.91
Conf. Int.	0.93	0.88	0.90	0.85	0.86	0.90
Alpha (5%)	0.95	0.91	0.93	0.90	0.91	0.93
Lambda1	0.78	0.67	0.68	0.66	0.67	0.68
Lambda2	0.94	0.90	0.91	0.88	0.89	0.91

Tab. 6. Reliability indicators for employee panel operations registration view

	Attractiveness	Perspicuity	Efficiency	Dependability	Stimulation	Novelty
Alpha	0.96	0.90	0.91	0.89	0.91	0.93
Conf. Int.	0.95	0.88	0.90	0.87	0.90	0.92
Alpha (5%)	0.96	0.92	0.93	0.91	0.93	0.94
Lambda1	0.80	0.68	0.695	0.67	0.69	0.70
Lambda2	0.95	0.90	0.91	0.89	0.92	0.93

In the general operations registration view (see Table 4), Cronbach's alpha ranges from 0.84 to 0.94 across the six scales. Corresponding 5% confidence intervals (e.g. 0.90–0.95 for attractiveness) generally indicate acceptable internal consistency. Lambda1 values range from 0.63 to 0.78, while Lambda2 ranges from 0.85 to 0.93, exceeding Alpha in most instances. In the Personalised Operations Registration View (Tab. 5), Alpha estimates (0.88–0.94) remain comparable, although Stimulation exhibits slightly reduced Alpha (0.89), which is still bounded by a tight confidence interval of 0.86–0.90. Again, Lambda2 (0.90–0.94) surpasses Alpha in nearly all scales, indicating stable lower-bound reliability. In the Employee Panel Operations Registration View (Table 6), Alpha averages are slightly higher for certain scales (0.90–0.93), with Lambda2 (0.89–0.95) closely aligning with or modestly exceeding these values. These patterns are consistent with the idea that Lambda2 offers a more reliable estimate of true reliability (Callender & Osburn, 2005), although minor computational artefacts may occasionally cause reversals (Schmitt, 1996). Overall, all three views demonstrate sufficient reliability, as evidenced by alpha coefficients that exceed the frequently cited threshold of 0.7 (Cronbach, 1951; Bonett, 2002). The marginal variations in alpha or lambda² across the three views do not appear to be indicative of major misinterpretations of items, suggesting that each UEQ scale is measured coherently despite potential contextual differences among the interfaces.

4.4. Benchmark

The scale of the evaluated product was rigorously compared to comprehensive benchmark data provided by the UEQ authors. This comprises responses from 21,175 participants across 468 studies, covering various product domains such as business software, websites, web shops and social networks. This benchmarking approach allows the product's UX metrics to be assessed objectively by situating them within a well-established reference framework. Benchmark chart variants were generated for each case, presenting mean scale scores and incorporating 95% confidence intervals to serve as a measure of the precision of these estimates. Including confidence intervals allows for a more nuanced interpretation of the scores, as it indicates the statistical reliability and accuracy of the assigned benchmark categories. A narrower confidence interval implies greater consistency among respondents' ratings and hence more robust conclusions regarding the product's comparative performance can be drawn. Ultimately, this benchmarking strategy is invaluable for identifying the strengths of ERP systems designed for SME production environments and pinpointing potential areas for improvement.

Tab. 7. Benchmark results for general operations registration view

Scale	Mean	Comparisson to benchmark	Interpretation
Attractiveness	0.43	Bad	In the range of the 25% worst results
Perspicuity	1.00	Below Average	50% of results better, 25% of results worse
Efficiency	0.95	Below Average	50% of results better, 25% of results worse
Dependability	0.84	Below Average	50% of results better, 25% of results worse
Stimulation	-0.13	Bad	In the range of the 25% worst results
Novelty	-0.54	Bad	In the range of the 25% worst results

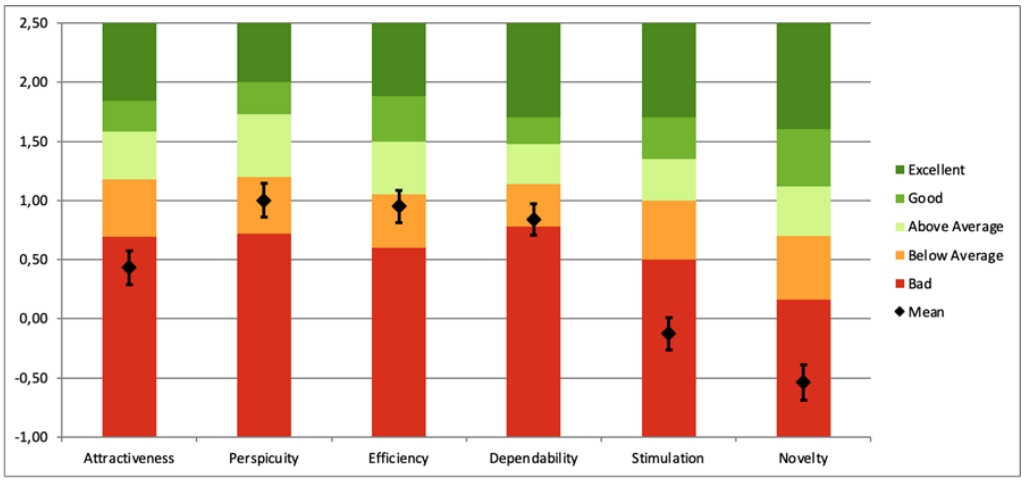


Fig. 13. Benchmark results chart for general operations registration view

Tab. 8. Benchmark results for personalized operations registration view

Scale	Mean	Comparisson to benchmark	Interpretation
Attractiveness	0.42	Bad	In the range of the 25% worst results
Perspicuity	0.77	Below Average	50% of results better, 25% of results worse
Efficiency	0.81	Below Average	50% of results better, 25% of results worse
Dependability	0.78	Below Average	50% of results better, 25% of results worse
Stimulation	0.01	Bad	In the range of the 25% worst results
Novelty	-0.19	Bad	In the range of the 25% worst results

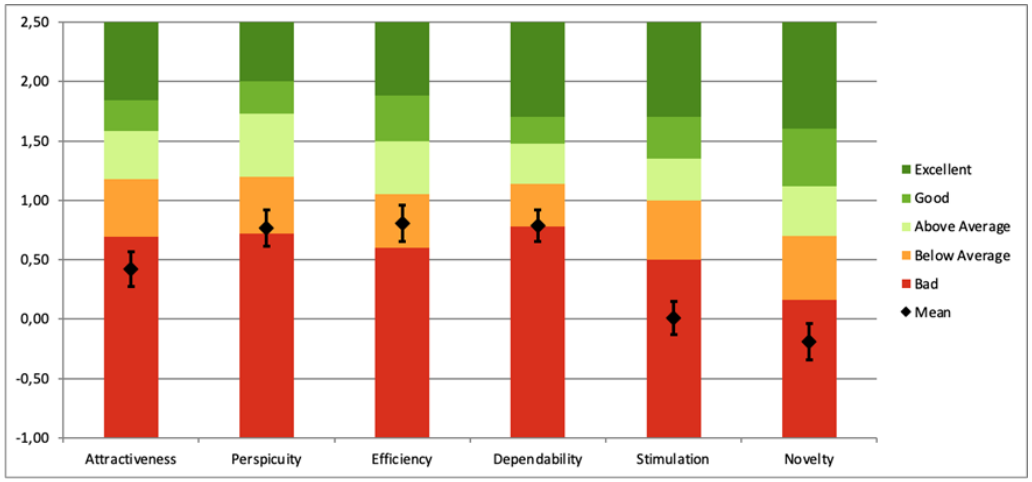


Fig. 14. Benchmark results chart for personalized operations registration view

Tab. 9. Benchmark results for employee panel operations

Scale	Mean	Comparisson to benchmark	Interpretation
Attractiveness	0.61	Bad	In the range of the 25% worst results
Perspicuity	0.59	Bad	In the range of the 25% worst results
Efficiency	0.72	Below Average	50% of results better, 25% of results worse
Dependability	0.64	Bad	In the range of the 25% worst results
Stimulation	0.52	Below Average	50% of results better, 25% of results worse
Novelty	0.52	Below Average	50% of results better, 25% of results worse

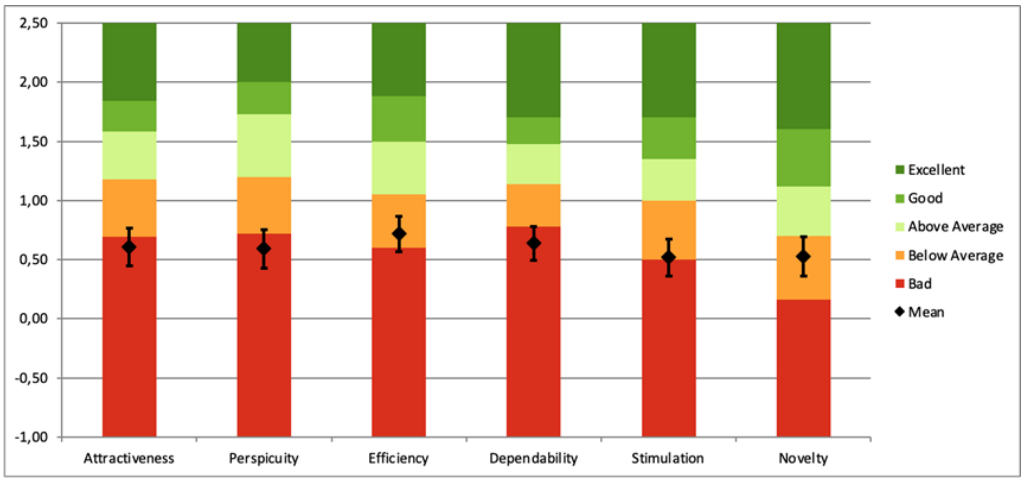


Fig. 15. Benchmark results chart for employee panel operations view

Benchmark analyses of the three interface variants — the General Operations Registration View (Table 7 and Figure 13), the Personalised Operations Registration View (Table 8 and Figure 14) and the Employee Panel Operations Registration View (Table 9 and Figure 15) — reveal consistently low attractiveness scores. This places them all in the bottom 25% of benchmark comparisons (i.e. the 'bad' category). More specifically, the General Operations Registration View (Attractiveness = –0.54) achieves a mean Perspicuity score of 1.00, yet still rates as 'Below Average', indicating that, while users find this layout somewhat comprehensible relative to other dimensions, it is weaker than half of the products in the reference dataset. By contrast, the Personalised Operations Registration View (Attractiveness = 0.42) achieves an Efficiency mean of 0.81, also "Below Average", suggesting moderate performance but still trailing 50% of benchmarked systems. Similarly, the Employee Panel Operations Registration View exhibits an Efficiency mean of 0.72 (also 'Below Average') and a Stimulation mean of 0.24, emphasising persistent deficiencies in perceived motivational appeal. Across all three interfaces, dependability hovers in the 'below average' zone (e.g. 0.84 in general, 0.81 in personalised and 0.72 in employee panel), pointing to concerns about the perceived control or predictability of interactions. Notably, novelty scores consistently negatively, reaching –0.54 in the general view, –0.19 in the personalised view and 0.52 in the employee panel, implying that users see little innovation or creative flair in these designs apart from a modest improvement in the latter. Collectively, these benchmark positions highlight the need for enhancements in visual appeal, stimulating design features and reliability across all three UI configurations. From a production engineering standpoint, such benchmarking results highlight opportunities to refine ERP interface design to better align with industry-leading solutions, thereby improving operational workflows and user satisfaction in SME manufacturing environments.

5. DISCUSSION

The empirical findings from our UEQ analysis provide valuable insights into the multifaceted nature of UX in ERP systems designed for SME production environments. In line with previous studies (Laugwitz et al., 2008; CK & Setiawan, 2021), the current analysis shows that the pragmatic dimensions — namely Perspicuity, Efficiency and Dependability — consistently achieve positive mean scores exceeding +0.8, indicating robust functional performance. Conversely, the hedonic dimensions, Stimulation and Novelty, yield neutral to negative mean values (e.g. Novelty at approximately −0.54 in the General Operations Registration View),

suggesting that, while the system effectively facilitates task-related functions, it lacks innovative, engaging features to inspire user interest. These findings highlight the ongoing challenge of balancing operational efficiency with an engaging, forward-thinking design aesthetic.

Benchmark comparisons utilising a comprehensive dataset of 21,175 respondents from 468 studies further position the evaluated product in the lower quartile for key dimensions such as attractiveness and hedonic quality across all interface variants. This discrepancy highlights the ongoing challenge of balancing operational efficiency with an appealing design. Furthermore, the narrow 95% confidence intervals — indicative of the high precision of our mean estimates — confirm the stability of our measurements despite some variability in hedonic ratings. Overall, the differential performance between pragmatic and hedonic aspects implies that, while current ERP interfaces satisfy basic operational requirements, there is significant scope for design enhancements aimed at elevating user engagement and satisfaction. The observed trends corroborate previous research and illuminate context-specific nuances that merit further investigation, particularly in reconciling functional efficacy with visual and experiential innovation. From a production engineering standpoint, the implications of these findings are profound. The consistently low benchmark scores for attractiveness and hedonic dimensions across all evaluated registration views highlight the critical need for interface redesign strategies that integrate advanced, user-centred design principles. Enhancements such as modernised graphics, interactive dashboard elements and adaptive navigation features could substantially improve the system's visual appeal and engagement levels, thereby addressing the observed shortcomings. Furthermore, the high internal consistency observed in pragmatic scales indicates that improvements in usability and system clarity could directly enhance operational performance by reducing user errors and optimising task completion times. These insights advocate a dual-focused approach combining functional optimisation with aesthetic refinement to ensure that ERP systems not only streamline production processes, but also foster a more stimulating user experience (UX). Collectively, the practical implications derived from this study provide ERP system designers and managers with actionable guidelines, reinforcing the idea that investing in refined UI/UX elements is critical for achieving a competitive advantage and operational efficiency in production-focused SMEs.

6. CONCLUSIONS

An empirical analysis of UEQ data has unequivocally demonstrated a significant difference between the pragmatic and hedonic dimensions of ERP interfaces used in ICT-driven production environments. In particular, our results show that the pragmatic quality aspects — namely Perspicuity, Efficiency and Dependability — regularly exceed the +0.8 threshold. This highlights the systems' robust functional performance and the high degree of operational clarity and control that users perceive. Conversely, the hedonic dimensions, specifically Stimulation and Novelty, consistently receive neutral to negative evaluations, indicating deficiencies in features related to user engagement and innovative appeal. Benchmark comparisons utilising an extensive dataset of over 21,000 responses across 468 studies position these ERP modules within the lower quartile for dimensions such as attractiveness and hedonic quality. This highlights a significant gap in aesthetic and experiential design relative to industry standards. While the data confirm that core operational functionalities are effectively supported, they also reveal considerable scope for redesigning the interface to enhance visual appeal and foster a more engaging user experience (UX). Furthermore, integrating precise statistical techniques, such as calculating 95% confidence intervals and using reliability indices like Cronbach's alpha and Guttman's lambda^2, ensures the observed trends are robust and the mean estimates are stable across various user interface views. In summary, integrating the latest UX trends, underpinned by an interdisciplinary research approach, facilitates the creation of IT systems that are functionally sound and user-centric. Investment in UX development, anchored in solid scientific foundations and enriched by methodologies from production and mechanical engineering, constitutes an investment in people and yields enduring socio-economic benefits. However, certain limitations must be acknowledged, including potential subjectivity arising from self-reported data and the predominantly academic nature of the sample. Future research should incorporate a larger proportion of production employees and longitudinal designs to capture the sustained impact of iterative interface improvements. Ultimately, these findings offer ERP system designers and managers actionable guidelines, suggesting that a dual-focused approach optimising both pragmatic functionality and hedonic appeal is critical for achieving operational excellence and sustained competitive advantage in production-oriented SMEs.

Conflicts of Interest

The authors declare no conflict of interest.

REFERENCES

Alves, J., Lima, T. M., & Gaspar, P. (2023). Is Industry 5.0 a human-centred approach? A systematic review. *Processes*, *11*(1), 193. https://doi.org/10.3390/pr11010193

Antonowicz, M., & Zaborowski, H. (2023). Bezpapierowe łańcuchy dostaw – rola i znaczenie. *Ekonomika I Organizacja Logistyki*, *8*(2), 121–141. https://doi.org/10.22630/EIOL.2023.8.2.16

Auksztol, J., Balwierz, P., & Chomuszko, M. (2015). *SAP. Zrozumieć system ERP*. Wydawnictwo Naukowe PWN.

Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics*, *27*(4), 335–340. https://doi.org/10.3102/10769986027004335

Callender, J., & Osburn, H. (2005). An empirical comparison of coefficient alpha, Guttman's Lambda-2, and MSPLIT maximized split-half reliability estimates. *Journal of Educational Measurement*, *16*(2), 89–99. https://doi.org/10.1111/j.1745-3984.1979.tb00090.x

Chomuszko, M. (2016). *System ERP – dobre praktyki wdrożeń*. Wydawnictwo Naukowe PWN.

CK, A. F. B., & Setiawan, J. (2021). Analysis of user experience resource planning with user experience questionnaire framework (case study: Universitas Multimedia Nusantara). *Journal of Multidisciplinary Issues*, *1*(2), 42–61. https://doi.org/10.53748/jmis.v1i2.21

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Fokczyński, W. (2023). Wdrożenie systemu ERP jako decyzja o znaczeniu strategicznym w przedsiębiorstwie z sektora MŚP w Polsce. *Informatyka w Biznesie*, 41-56. http://dx.doi.org/10.15611/2023.81.7.04

Gross-Gołacka, E., & Ożga, J. (2023). Kapitał intelektualny w organizacji wobec idei zrównoważonego rozwoju w dobie Społeczeństwa 5.0. *Prakseologia i Zarządzanie*, 1, 59-71.

Hamera, A. (2024). Rola doświadczeń użytkownika w erze Przemysłu 5.0. *Zarządzanie Przedsiębiorstwem: Enterprise Management*, *27*(2), 7–15.

Hassenzahl, M., & Tractinsky, N. (2006). User experience – A research agenda. *Behaviour & Information Technology*, *25*(2), 91–97. http://doi.org/10.1080/01449290500330331

International Organization for Standardization. (2010). *ISO 9241-210:2010 Ergonomics of human-system interaction: Human-centred design for interactive systems*. https://www.iso.org/standard/52075.html

Knosala, R. (2023). *Inżynieria zarządzania. Cyfryzacja produkcji. Aktualności badawcze 5*. Polskie Wydawnictwo Ekonomiczne.

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In A. Holzinger (Ed.), *HCI and Usability for Education and Work* (Vol. 5298, pp. 63–76). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-89350-9_6

Longo, F., Padovano, A., & Umbrello, S. (2020). Value-oriented and ethical technology engineering in Industry 5.0: A human-centric perspective for the design of the factory of the future. *Applied Sciences*, *10*(12), 4182. http://dx.doi.org/10.3390/app10124182

Matuszek, J., & Kurczyk, D. (2017). Projektowanie procesów i systemów produkcyjnych z wykorzystaniem technologii komputerowej wirtualizacji. *Innowacje w zarządzaniu i inżynierii produkcji*, *2*(12), 741–752.

Mleczko, J. (2015). Integration of CAD/PDM and ERP systems in practice. *Applied Mechanics and Materials*, *791*, 26–33. https://doi.org/10.4028/www.scientific.net/AMM.791.26

Mleczko, J. (2016). The paperless factory as a result of integration between CAD/PDM and ERP systems. In D. Plinta (Eds.), *Advanced industrial engineering: Industry 4.0: monograph* (pp. 57–72). Wydawnictwo Fundacji Centrum Nowych Technologii.

Munandar, F., & Santoso, H. B. (2025). User interface improvement by evaluating usability and user experience: Case study of Indonesia's Government Financial Management Information System. *The Indonesian Journal of Computer Science*, *14*(1), 249-266. https://doi.org/10.33022/ijcs.v14i1.4630

Norman, D., & Nielsen, J. (1998, August 8). The definition of user experience. http://www.nngroup.com/about/userexperience.html

Nugraha, I., & Fatwanto, A. (2021). User experience design practices in industry (case study from Indonesian information technology companies). *Electronics, Informatics, and Vocational Education*, *6*(1), 49-60. https://doi.org/10.21831/elinvo.v6i1.40958

Parys, T. (2005). Bariery wdrożeniowe systemu zintegrowanego klasy ERP i ich postrzeganie przez użytkowników. In M. Rószkiewicz, E. Wędrowska (Eds.), *Informacja w Społeczeństwie XXI wieku (Monografie i Opracowania No. 540)*.

Parys, T. (2010). Wdrożenie zintegrowanego systemu informatycznego klasy ERP jako zmiana organizacyjna. *Informacja w Społeczeństwie XXI wieku*.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350-353. https://doi.org/10.1037/1040-3590.8.4.350

Schrepp, M. (2015). User experience questionnaire handbook. *SAP Research*. http://dx.doi.org/10.13140/RG.2.1.2815.0245

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Construction of a benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence*, *4*(4), 40–44. http://doi.org/10.9781/ijimai.2017.445

Simundic, A.-M.(2008). Confidence interval. *Biochemia Medica*, *18*(2), 154–161. https://doi.org/10.11613/BM.2008.015

Winter, J. (2025, January 22). Success factors for a paperless ecosystem in manufacturing. Jeff Winter Insights. https://www.jeffwinterinsights.com/insights/success-factors-for-a-paperless-ecosystem-in-manufacturing