Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Usability of Intensive Care Ventilators: A Heuristic Evaluation by Expert Clinicians

Soheyla Ahmadi Charkhabi
Shiraz University of Medical Sciences
Hossein Parsaei
Shiraz University of Medical Sciences
Seyed Masoom Masoompour
Shiraz University of Medical Sciences
Hamid Azizzadeh
Pooyandegan Rah Saadat Co
Jafar Hassanzadeh
Shiraz University of Medical Sciences
Abolfazl Massoudi Rad
Shiraz University of Medical Sciences
Seyed Meysam Mortazavi
Yasuj University of Medical Sciences
Alireza Choobineh
alrchoobin@sums.ac.ir

Shiraz University of Medical Sciences

Article

Keywords: Medical device interface, Usability, Mechanical ventilators, Heuristic evaluation, Safety-critical systems, intensive care unit (ICU)

Posted Date: June 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9671815/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. H.A. is employed by a medical device company that manufactures one of the ventilator models evaluated in this study. To mitigate potential bias, this author was not involved in data collection, usability evaluations, severity rating, data analysis, or

interpretation of the results related to the evaluated devices. All evaluations and analyses were conducted independently by the remaining authors. The authors confirm that this employment relationship did not influence the study design, data collection, analysis, or reporting of the findings.

Abstract

Mechanical ventilators in intensive care units (ICUs) are safety-critical interactive systems in which usability failures can compromise patient safety. Evidence shows many ventilator-related adverse events arise from breakdowns in human–computer interaction rather than device malfunction. This study evaluated the usability of seven ICU ventilators in real clinical settings. Twenty-one clinicians completed 42 in-situ heuristic evaluation sessions after structured training. Identified usability problems were consolidated and severity-rated on a five-point scale. Quantitative analyses examined violation patterns and severity distributions, while reflexive thematic analysis explored underlying human-factors mechanisms of high severity issues.

Heuristic violations were most commonly involving poor alignment with clinical workflows, weak error prevention, and limited visibility of system state. Severity profiles differed across devices, and greater evaluator experience was linked to lower severity ratings. Four recurring interaction breakdowns emerged: perceptual limitations, excessive cognitive load, inadequate error safeguards, and workflow misalignment, underscoring the need for cognitively aligned interface design.

1. Introduction

Mechanical ventilators are essential life-support technologies in intensive care units (ICUs), central to the management of acute respiratory failure. Their critical role became especially visible during the COVID-19 pandemic, which highlighted not only global equipment shortages but also challenges in ensuring safe and effective use under extreme clinical pressure [1, 2]. Beyond emergency contexts, the global burden of chronic respiratory diseases increased substantially between 1990 and 2017, with continued growth in mortality and disability-adjusted life years. These trends reinforce the status of ventilators as permanent, safety-critical components of healthcare systems [3, 4]. Nevertheless, evidence indicates that many ventilator-related adverse events stem not from technical malfunction but from failures in human–machine interaction within the high-pressure ICU environment [5, 6].

Advances in ventilator technology have produced increasingly sophisticated systems with numerous ventilation modes, adjustable parameters, and alarm functions, substantially increasing system complexity [7–9]. While enhancing clinical capabilities, these developments also challenge information architecture and increase cognitive workload, particularly in ICUs [10, 11]. Modern interfaces often present large volumes of information without clear prioritization, exceeding clinicians' processing capacity in high-stress settings and contributing to cognitive overload and alarm fatigue [7, 9, 12].

Empirical studies show that even experienced clinicians struggle with critical tasks due to deep menu hierarchies, inconsistent terminology, and non-intuitive interfaces [7, 8, 10]. These usability flaws can cause errors, misinterpretation of system feedback, and delays in time-sensitive interventions, directly compromising patient safety [11, 12]. Notably, prior research indicates that device simplicity alone does not ensure safe use if feedback mechanisms and interface logic are misaligned with clinicians' mental models and workflows [13].

Although ventilator usability has been studied, much research occurs in simulated or laboratory settings, failing to capture the physical, cognitive, and organizational complexity of real-world ICUs [14, 15]. Many evaluations involve human factors or user experience (UX) specialists without clinical expertise and focus on isolated tasks rather than the ventilator as an integrated system within daily workflows [8, 16, 17]. Retrospective analyses of incident reports, while useful for classifying errors, act as lagging indicators and cannot proactively identify high-severity latent design flaws [5, 6].

Some studies indicate that ventilator development has prioritized technical functionality and regulatory compliance over alignment with clinicians' cognitive constraints and workflows, allowing critical usability flaws to persist in approved systems. Thus, a gap remains in research involving clinically trained professionals as dual experts, integrating domain knowledge with usability principles to assess ventilators holistically in real-world use [16, 18].

Evidence indicates that many problems reported by clinicians' stem from usability deficiencies rather than technical faults, and heuristic evaluation can proactively uncover these vulnerabilities [19]. Involving physicians and nurses as evaluators reveals mismatches between interface design, workflows, and environmental constraints often missed in laboratory or engineer-led assessments [1, 20]. This approach supports regulatory frameworks emphasizing inherent safety through design rather than reliance on downstream measures like training or labeling [21, 22].

Given the safety-critical role of mechanical ventilators, usability is a key determinant of safe and effective care. Although usability deficiencies can directly threaten patient safety, no prior study has systematically evaluated ICU ventilators using heuristic methods with end users in real clinical settings. This study addresses this gap by assessing seven commonly used ICU ventilators through the 14-principle heuristic by Zhang et al. [16]. applied by physicians and nurses. Beyond identifying usability problems, it examines how professional role and clinical experience affect risk perception and severity ratings, and uses thematic analysis to explore high-severity issues, providing evidence to inform the design and procurement of safer, user-centered ventilator systems.

2. Materials and Methods

2.1 Study Design and Framework

This descriptive-analytical study conducted a comprehensive usability assessment of ICU ventilators using qualitative expert data derived from heuristic evaluation method. The approach was used to identify usability problems, determine the severity of heuristic violations for each device, and translate findings into actionable design recommendations. Heuristic evaluation is a well-established usability engineering method shown to effectively identify design violations associated with medical error with relatively low resource requirements. Its ecological validity is strengthened when clinicians perform evaluations within the real context of use. For these reasons, the 14- heuristic evaluation introduced by

Zhang et al. [16] was selected as an efficient and clinically relevant approach for evaluating ventilator usability.

2.2 Heuristic Evaluation

2.2.1 Ventilator Selection

Following an inventory of ventilator models conducted across ten educational hospitals affiliated with Shiraz University of Medical Sciences in November 2024, seven contemporary ICU ventilator models (Drager Savina 300 (DS), Drager Evita XL (DEX), Drager Evita 4 (DEV), Hamilton C2 (HC), Maquet Servo I (MSI), Respina P1 (R), and EDP- TS (E) were selected for evaluation. Selection was based on three criteria: exclusive use in ICUs, high frequency of clinical use, and representation of current-generation technology within these hospitals.

2.2.2 Participants

A purposive sample of 21 expert clinicians (16 nurses and 5 physicians) was recruited from ICU staff across the hospitals. Participants were identified through peer referral based on extensive ventilator experience and strong professional performance. Of the 33 clinicians approached, 21 agreed to participate (response rate: 63.6%), with non-participation mainly due to ICU workload constraints. In Iran, respiratory therapy is not recognized as a distinct profession; therefore, ICU nurses and physicians (including intensivists and pulmonary specialists) represent the primary ventilator users.

2.2.3 Evaluator Training

To standardize the evaluation process, a one-hour instructional video was developed by the research team. This video trained the clinicians on the principles of heuristic evaluation, using the 14-principle framework for medical devices adapted by Zhang et al. [16]. In this training video, each heuristic principle was defined, followed by illustrative examples of violations. To prevent assessment bias, these examples were drawn specifically from ventilator models other than those selected for the current study. An asynchronous training approach was adopted, with the video distributed to participants via secure messaging platforms, allowing them to complete the training at their convenience. The strategic objective of this training was to leverage the participants' deep understanding of the clinical context. As end-users, these clinicians are acutely aware of how device challenges directly impact workflow and treatment efficacy. By equipping them with formal usability assessment skills, this study aimed to elevate them to the status of dual-domain experts, evaluators possessing both high-level domain expertise and the necessary methodological knowledge to conduct valid heuristic inspections.

2.2.4 Evaluation Procedure

A total of 42 in situ evaluation sessions were conducted at the participants’ respective hospitals using a standardized protocol. Each session began with a 15 to 20-minute briefing led by the researcher to review heuristic principles and clarify the procedure, followed by completion of demographic and occupational form. Then a written consent form which participants signed were obtained. Clinicians then

evaluated a ventilator model with which they had prior experience during a 60 to 90-minute structured walkthrough using a lung simulator. Participants identified usability problems and the corresponding violated heuristic. All sessions were audio-recorded with participant's consent. Each of the seven devices was evaluated by a four to nine clinicians.

2.3 Data Analysis

2.3.1 Analysis of Usability Problems and Severity

Audio recordings were transcribed verbatim, and identified usability problems were compiled and synthesized in MAXQDA 2018. Duplicate issues within each device were merged to generate a final list of unique problems, which was reviewed and validated by the research team in three in-person sessions, including on-site clarification of ambiguous cases.

The finalized problem list was returned to the original evaluators for confirmation and severity rating. Each evaluator received a coded version of their entries to verify accurate interpretation and merging; corrections were provided in writing when needed. If the interpretation was accurate, the evaluator confirmed it by signing the usability problem list. Evaluators then rated problem severity on a 5-point scale (0–4). Inter-rater reliability for each device was assessed using a two-way mixed-effects Intraclass Correlation Coefficient (ICC) for absolute agreement, accommodating variable numbers of raters. Mean severity scores were calculated and classified into four levels include: cosmetic (< 1.5), minor (1.5–2.4), major (2.5–3.4), and catastrophic (≥ 3.5) based on Zhang et al. [16].

2.3.2 Thematic Analysis of Critical Issues

Following severity classification, a targeted reflexive Thematic Analysis (TA) was conducted exclusively on usability problems rated as major or catastrophic, representing the most significant threats to use errors, patient safety, and treatment efficacy. Adhering to Clarke and Braun's six-phase framework [23], the analysis transcended semantic description to interpret underlying ergonomic failures. Initially, the dataset was immersed through repeated reading to identify patterns, and initial codes were generated in MAXQDA to capture both explicit user complaints (e.g., illegible fonts) and latent human factors concepts (e.g., cognitive load). Codes were then clustered into provisional themes based on shared conceptual meaning rather than device components.

Credibility and trustworthiness were ensured through iterative team discussions during phases three to five of TA framework, refining codes, investigating theme boundaries, and merging overlapping sub-themes (e.g., combining standby logic with adaptive automation failures). For example, the 'rigid navigation' sub-theme, initially classified under operational workflow (theme 4), was reassigned to cognitive load (theme 2) because deep menu structures imposed demands on working memory and decision-making, rather than a physical workflow impediment. Throughout, the 16-item checklist proposed by Ahmed et al. [24] was applied to ensure adherence to TA rigor. Final themes were defined using formal human factors terminology to provide actionable insights for design teams on safety-critical interface areas.

2.3.3 Statistical Analysis

Statistical analyses were conducted using IBM SPSS Statistics version 26. Descriptive statistics summarized participant characteristics and the distribution of usability problems. A Chi-Square Goodness-of-Fit test assessed the distribution of violations across the 14 heuristic principles. Associations between ventilator models and problem severity categories were examined using Fisher's Exact Test with Monte Carlo simulation (10,000 samples). Spearman's rank-order correlation evaluated relationships between evaluator team characteristics and evaluation outcomes. Inter-rater reliability of severity ratings was assessed using a two-way mixed-effects Intraclass Correlation Coefficient (ICC) for absolute agreement. Statistical significance was set at $\alpha = 0.05$.

2.4 Ethical Considerations

This study was approved by the Ethics Committee of Shiraz University of Medical Sciences (IR.SUMS.SCHEANUT.REC.1402.098). All research was performed in accordance with relevant guidelines and regulations, and in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants involved in the study.

3. Results

3.1 Participant Characteristics

A total of 21 expert clinicians (16 nurses, 5 physicians) served as heuristic evaluators. Table 1 displays demographic details of the participants.

Table 1
Demographic and Professional Characteristics of Study Participants

Characteristic	Heuristic Evaluators (N = 21)
Specialty	**n (%)**
Physician	5 (23.8%)
ICU Nurse	16 (76.2%)
Gender	
Male	13 (61.9%)
Female	8 (38.1%)
Education Level	
Bachelor's Degree	13 (61.9%)
Master's or Ph.D.	3 (14.3%)
Anesthesiology specialist	4 (19%)
Pulmonary specialist	1 (4.7%)
	Mean (SD)
Age (years)	53.5 (10.64)
Experience with Ventilators (years)	13.45 (8.10)
Tenure (years)	18.45 (10.13)

[Table 1]

3.2 Heuristic Usability Violations

Table 2 presents sample findings from the heuristic evaluations, illustrating usability problems identified across the ventilator models. Each problem received a severity rating, with mean scores determining its classification, and was linked to one or more violated heuristic principles. Overall, 344 unique usability problems and 1177 heuristic violations were identified, with detailed results provided in the following sections.

Table 2
the usability problems, its severity and heuristic violating for some samples of the usability problem of all seven devices.

Usability problem description	**Heuristic violated**	**Mean (SD) severity rating***
The ventilator does not automatically return to the prior ventilation mode after apnea resolution, risking prolonged inappropriate therapy.	Consistency, Match, Feedback, Error, Control	3.7 (0.4)
Alarm messages lack clear explanations, causes, and actionable recovery guidance.	Feedback, Message, Language, Control, Help	2.6 (0.8)
Low contrast, small fonts, and poorly distinguishable icons on both screen and hardware reduce legibility (white text on blue background, white arrow on white background for tubing connection).	Consistency, Visibility, Error	2.8 (0.9)
Inappropriate default value for the respiratory rate in the apnea backup settings (default is 6).	visibility, match, feedback, error, memory, help	3.2 (0.9)
Components can be assembled incorrectly due to absent orientation cues and lack of mechanical constraints.	visibility, match, feedback, error, closure, control	2.6 (0.4)
Patient settings cannot be restored quickly after power loss	Match, Flexibility, Error, Closure, Undo, Control	3.2 (0.4)
Users must exit current screens before accessing others, increasing interaction cost.	Consistency, Flexibility, Control	2.6 (0.8)
Frequent oxygen sensor failures undermine system trust.	Match, Error, Control	3.5 (0.5)
Parameter changes are lost if mode confirmation is missed, with no feedback to the user.	Match, Memory, Feedback, Error, Closure, Control	2.7 (0.4)
Critical parameter changes (e.g., I: E inversion) are not made sufficiently salient.	Consistency, Visibility, Feedback, Error	3.2 (0.4)
Failure messages display codes without explanation or recovery instructions.	Visibility, Message, Language, Help	3.0 (0.7)
Lengthy pre-use checks promote unsafe workarounds and skipped safety procedures.	Match, Minimalist, Flexibility, Error, Control	3 (0.8)
The device does not exit standby automatically upon sensing patient effort.	Match, Error, Control	3.6 (0.4)

**The mean severity score for each problem was min = 0 to max = 4*

Usability problem description	Heuristic violated	Mean (SD) severity rating*
Related settings (e.g., apnea backup time) are not collocated, increasing memory burden.	Visibility, Match, Error	2.5 (0.7)
Harsh, poorly chosen color schemes contribute to eye strain.	Visibility, Minimalist	2.5 (0.9)
Users can set physiologically unsafe alarm limits (e.g., RR = 0).	Match, Error	3.3 (0.7)
Nebulization provides no confirmation of duration, flow, or proper operation.	visibility, Feedback, Closure, Control, Help and documentation	2.6 (0.4)
Users can disable essential sensors (e.g., flow sensor), degrading monitoring and therapy.	Error	3.0 (0.5)
Absence of common ventilation modes (e.g., BiPAP, ASV) limits clinical flexibility.	Match, Flexibility, Control	3.5 (0.4)
Informational messages are poorly positioned, visually indistinct, and fail to capture attention.	Visibility, Match, Minimalist, Feedback, Message, Error	3.6 (0.5)
Disabling the exhalation flow sensor produces no system feedback.	Match, Feedback, Error, Control, Help	3.6 (0.5)
**The mean severity score for each problem was min = 0 to max = 4*		

[Table 2]

3.3 Distribution and Frequency of Heuristic Violations

A detailed frequency analysis, summarized in Table 3, identified a distinct hierarchy of problem categories. To determine if usability problems were concentrated in specific design areas, the distribution of all 1177 identified heuristic violations was analyzed across the 14 heuristic principles.

Table 3
Frequency and distribution of heuristic violations across all ventilators (N = 1177*)

Rank	Fourteen usability heuristics	Frequency (n)	Percent of Total (%)	Cumulative Percent (%)
1	Match between system and world [match]	219	18.6%	18.6%
2	Prevent error [error]	129	11.0%	29.6%
3	Visibility of system state [visibility]	128	10.9%	40.4%
4	User in control [control]	116	9.9%	50.3%
5	Informative Feedback [feedback]	104	8.8%	59.1%
6	Consistency and standards [Consistency]	103	8.8%	67.9%
7	Help and documentation [help]	80	6.8%	74.7%
8	Flexibility and efficiency [Flexibility]	78	6.6%	**81.3%**
9	Minimalist	64	5.4%	86.7%
10	Minimize memory load [memory]	48	4.1%	90.8%
11	Use users' language [language]	36	3.1%	93.9%
12	Good error messages [messages]	35	3.0%	96.9%
13	Clear closure [closure]	29	2.5%	99.3%
14	Reversible actions [undo]	8	0.7%	100.0%
Total		**1177**	**100.0%**	

**total number of heuristic violations is more than usability problems because one problem may violate multiple heuristics*

The single most violated principle was Match between system and world, which accounted for 18.6% of all identified violations. This was followed by Error Prevention (11.0%) and Visibility of System State (10.9%). Together, these top three heuristic violations constituted over 40% of all usability issues found in the study. A Pareto analysis of the distribution indicated that the eight heuristic principles accounted for 81.3% of all identified violations. A Chi-Square Goodness-of-Fit test revealed a statistically significant deviation from a uniform distribution, χ^2 (13, N = 1177) = 466.567, $p < 0.001$. This confirms that certain categories of usability violations occurred significantly more frequently than others.

[Table 3]

3.4 Comparative Analysis of Heuristic Violation Profiles Across Devices

The total number of identified usability problems varied considerably between the ventilator's models (Table 4). A greater number (N = 72) of usability issues were identified for the R device. These usability problems were associated with a total of 250 specific heuristic violations (as a single usability problem often violated multiple heuristic principles or was flagged by multiple evaluators), while Device MSI had the lowest number (N = 38) of unique usability problems.

Table 4
Frequency and Proportional Distribution of Usability Problems by Severity Category

Ventilator Model	Total Problems (N)	Catastrophic (n)	Major (n)	Minor (n)	Cosmetic (n)	Critical Problem Percentage *
R	72	3	52	17	0	76.4%
DEX	42	3	21	15	3	57.1%
HC	52	2	17	30	3	36.5%
DEV	59	0	21	32	6	35.6%
E	40	0	14	25	1	35.0%
MSI	38	1	11	24	2	31.6%
DS	41	0	10	30	1	24.4%
Total	344	9	146	173	16	

**Percentage of Severe Issues calculated as ((Major + Catastrophic problems) / Total problems)* * 100*

[Table 4**]**

Beyond the total counts, a proportional analysis highlighted that the types of usability challenges were distinct to each device. For example, the device R's profile was dominated by violations of the Match between system and world and Error Prevention principles, suggesting that its primary flaws are related to unintuitive design and a failure to prevent user mistakes. In contrast, Device DS's profile was uniquely characterized by a high proportion of violations related to Visibility of system state, indicating its main challenges are in providing clear system feedback. (Fig. 1).

[figure 1**]**

To determine if the profile of usability problem severity differed across the seven ventilator models, a statistical analysis was performed. Using a Monte Carlo simulation of Fisher's Exact Test, the results

indicated a statistically significant association between the ventilator model and the severity category of its identified problems ($p < 0.001$).

As summarized in Table 4 and illustrated in Fig. 2, the analysis of severity profiles revealed a clear risk hierarchy among the devices. Device R had 76.4% usability flaws which classified as high-severity (major or catastrophic). In contrast, device DS was a lowest-risk model, where a similar proportion of issues (73.2%) were classified in the less critical minor category. The remaining devices formed a middle category, with the proportion of high-severity problems ranging from 31.6% to 36.5%.

[figure 2]

3.5 Influence of Evaluator Team Characteristics on Evaluation Outcomes

To assess whether the composition of expert evaluation teams influenced device-level findings, a Spearman's rank-order correlation was conducted at the device level (N = 7 devices; Table 5). This analysis examined the relationship between two team characteristics, proportion of physicians and mean years of ventilator work experience, and two outcomes: total number of usability issues and proportion of high-severity problems.

Table 5
Summary of Heuristic Evaluation Results by Device

Device	Proportion of Severe Issues *	Percent Physicians of evaluators	Mean Team Experience with ventilator (years)
DS	0.24 (10/41)	20% (1 of 5)	21.80 (11.08)
DEV	0.35 (21/59)	17% (1 of 6)	17.16 (11.86)
DEX	0.57 (24/42)	25% (1 of 4)	14.25 (4.02)
HC	0.36 (19/52)	17% (1 of 6)	11.33 (3.03)
R	0.76 (55/72)	22% (2 of 9)	12.55 (6.39)
E	0.35 (14/40)	20% (1 of 5)	15.10 (11.82)
MSI	0.31 (12/38)	33% (2 of 6)	14.83 (3.38)
**Proportion of Severe Issues calculated as (Major + Catastrophic problems) / Total problems.*			

No significant correlation was found between the proportion of physicians and either total issues ($r_s = -0.440$, $p = 0.323$) or high-severity problems ($r_s = 0.055$, $p = 0.907$). While average experience did not significantly affect total issues ($r_s = -0.396$, $p = 0.379$), it was strongly and negatively correlated with severity ratings ($r_s = -0.775$, $p = 0.041$), indicating that more experienced teams tended to assign lower severity scores. This suggests that teams with less experience rated issues as more severe, highlighting the influence of evaluator experience on perceived risk.

[Table 5]

3.6 Reliability of Problem Severity Ratings

To assess the consistency of the severity ratings assigned by the clinical evaluators, inter-rater reliability was calculated using an Intraclass Correlation Coefficient (ICC). Agreement was statistically greater than chance for five of the seven devices ($p < 0.001$), though ICC values were generally low (ranging from 0.109 to 0.181). For the remaining two devices (E and DS), the agreement was not statistically significant ($p > 0.05$) (Table 6).

Table 6
Inter-Rater Reliability of Severity Ratings by Ventilator Model

Ventilator Model	N (Evaluators)	ICC (Single Measures, Absolute Agreement) *	95% Confidence Interval	Significance (p-value)
DEX	4	0.181	0.046–0.354	< 0.001
HC	7	0.168	0.077–0.298	< 0.001
DEV	6	0.127	0.050–0.234	< 0.001
MSI	6	0.122	0.036–0.255	< 0.001
R	9	0.109	0.054–0.188	< 0.001
E	5	0.041	-0.038–0.166	0.172
DS	5	0.024	-0.040–0.127	0.250

**The table is sorted in descending order of ICC value. ICC values < 0.5 indicate poor agreement; values between 0.5 and 0.75 indicate moderate agreemen*

[Table 6]

3.7 Qualitative Analysis of High-Severity Usability Problems

Of all identified usability problems, 155 were rated as high severity (146 major and 9 catastrophic). Because these issues represent the most critical risks to patient safety and clinical performance, they were selected for in-depth thematic analysis. The thematic analysis identifies four comprehensive themes characterizing the usability of the ICU ventilators. Collectively, these themes point to a misalignment between the devices' engineering logic and the clinical realities of the ICU. The data suggests that the devices impose significant perceptual and cognitive demands on the operators, appears to cause deficiency essential safety resilience, and does not fully align with the physical and procedural context of critical care.

The following sections present the resulting themes and subthemes, illustrated with representative participant quotes, while Table 7 provides a structured overview of the thematic findings.

Table 7
Thematic analysis of usability challenges in the ICU ventilator, organized by theme, sub-theme, and specific failure categories.

Main Theme	Sub-theme	Code Categories (Clusters)	Key Example codes
1. Perceptual Deficits and Lack of System State Feedback	1.1 Visual Accessibility and Legibility Failures	• Illegible Typography and Sizing • Contrast, Color, and Glare Issues • Visual Clutter and Poor Hierarchy • Visibility in Context (Distance /Night) •Unnoticeable Alerts	- Font size too small for warnings. - decimal points unreadable. - Red text on blue background causes hard to see. - Screen cluttered without prioritization of vital signs. - Poor visibility during night shifts or from distance. - Critical alerts lack visual saliency.
	1.2 Invisible System Status	• Power and Battery Status Invisibility • Lack of Feedback Loops • Silent Parameter Changes • Unclear Configuration Status • Ambiguous Data Visualization	- Missing quantitative battery life indicator. - System provides no confirmation after self-tests. - Dependent parameters change silently without notification. - Uncertainty regarding sensor or nebulizer activation. - Graphs labeled by units rather than parameter names.
	1.3 Affordance and Signifier Failures	• False Static Affordance (Flat Design) • Ambiguous Iconography • Unclear Physical Control Affordance	- Interactive buttons resemble static text (flat design). - Standby icon is confusing or non-standard. - Physical buttons lack clear labels indicating function.

Main Theme	Sub-theme	Code Categories (Clusters)	Key Example codes
2. Cognitive load and Lack of System Transparency	2.1 Reliance on Error-Prone Memory	• Prospective Memory Burdens • Parameter Setting Recall • Manual Process Monitoring	- User must remember to turn off nebulizer manually. - User must memorize old settings to enter new ones. - Reliance on user vigilance rather than automation.
	2.2 Mental Model Mismatch and Terminology	• Terminology and Acronym Confusion • Mapping and Grouping Errors • Conceptual Mismatches • Calculation Burden • Obscure/Hidden Features	- Confusion between flow and L/min labels. - Illogical menu grouping (e.g., NIV hidden in settings). - Mental math required for Pressure Support calculation. - Essential features hidden in deep sub-menus.
	2.3 Rigid Navigation and High Interaction Cost	• Excessive Interaction Cost • Rigid/Linear Navigation Logic • Search and Discovery resistance • Lack of Direct Manipulation • Inefficiency in Critical Tasks	- Mandatory extra keystrokes (e.g., Exit before Enter). - Rigid sequences trap users in check cycles. - High visual search time to find basic parameters. - Unable to adjust parameters directly on main screen. - Emergency start-up requires too many steps.
	2.4 Unclear Logic and Hidden Dependencies	• Hidden Parameter Dependencies • Fragmentation of Related Functions	- Silent lockouts prevent changes without explanation. - Related apnea settings scattered across different screens.
3. Insufficient Error Prevention and Absence of Safety Constraints	3.1 defect of Error-Prevention Mechanisms	• Permissibility of Fatal Inputs	- System permits setting Respiratory Rate alarm limit to 0.

Main Theme	Sub-theme	Code Categories (Clusters)	Key Example codes
		• Unsafe Disabling of Components • Lack of Confirmation/Interlocks • Deficient Intelligent Constraints	- User can disable essential flow sensors without confirmation. - Immediate shutdown allowed without "Are you sure?" prompt. - Alarm limits allow nonsensical values (e.g., 100% leak).
	3.2 Automation Discrepancies and State Transitions	• Deficiencies in Automated State Transitions • Apnea Backup Failures • confusion of State Awareness	- Failure to auto-start ventilation upon patient connection. - Manual reset required after patient recovers from apnea. - Confusion regarding active mode vs. standby state. - Automation fails to adapt to patient effort
	3.3 Deficiencies in Alarm Logic and Error Recovery	• Cryptic/Non-Actionable Alarms • Poor Troubleshooting Support • System Resilience Failures • Alarm Logic and Fatigue Factors • Missing Maneuver Instructions	- Alarms displayed as technical error codes (e.g., Error 356). - No help/troubleshooting for alarms. - Loss of patient data after power cycle. - Premature/nuisance alarming causing desensitization. - Lack of guidance for hold maneuvers.
4. Operational Inefficiency and Contextual Incompatibility	4.1 Workflow Inefficiency and Clinical Inflexibility	• Clinical state Inflexibility • Mismatch with Patient Needs • Operational Friction/Delays • Nebulizer Workflow Mismatch	- Mandatory volume-control for apnea backup (inflexible). - Pediatric/Adult settings not clearly distinct. - Workflows require workarounds for standard care.

Main Theme	Sub-theme	Code Categories (Clusters)	Key Example codes
		• Inappropriate Default Settings	- Fixed nebulizer parameters (no time control). - Defaults do not match common clinical practice
	4.2 Physical features and Maintainability Challenges	• Maintainability Challenges • Hardware Fragility and Reliability • Physical Access and Hazards • Mobility Impediments • Environmental Stressors	- Complex cleaning protocols lead to non-compliance. - Frequent sensor breakage/failure. - Side-mounted tubing create clutter and obstruction. - Heavy compressor hinders intra-hospital transport. -Excessive noise generation.

[Table 7]

Theme 1: Perceptual Deficits and Lack of System State Feedback

This theme narrates failures at the sensory and perceptual stages of information processing. The interface frequently fails to bridge the gap between system state and user perception. As a result, the user unable to readily perceive the device's status or interpret available controls. Rather than providing transparent and at-a-glance status updates, the system appears to lack transparency, necessitating excessive visual search and interpretation. This theme comprises three subthemes that address different design-related dimensions of the problem in ventilator interfaces, which are described in detail below with the quotes of clinicians.

1.1. Visual Accessibility and Legibility Failures

Evaluators indicated that the interface design may not fully meet ergonomic standards and usability principals regarding legibility and contrast, particularly given the environmental constraints of an ICU (e.g., dim lighting during night shifts, viewing distance from the patient's bedside). Critical alphanumeric data were described as lacking enough visual hierarchy, requiring clinicians to approach the screen closely to verify parameters, which diverts attention from the patient.

The font is too small for me. As I'm getting older, my eyes get tired during night shifts. It's really hard to read these items with this black color and gray background. phrase like '100% O2' or the 'Hold' buttons on that background are barely visible. That thin font is just really tough to see. (Experience ICU supervisor nurse).

1.2. Invisible System Status

Beyond legibility, the system exhibits a lack of status visibility. Clinicians described the device as difficult to interpret regarding existing background states, such as battery indicator, power source, and sensor functionality. The absence of continuous feedback loops prevents the user from maintaining accurate situational awareness, leading to uncertainty about clinical condition of the patient.

It just shows the 'plugged-in' icon. No percentage or remaining time. If I want to transport a patient, I don't know if the battery is enough. I'm worried it might die mid-transfer and I'd have no way of knowing. (Senior intensive care expert).

1.3. Affordance and Signifier Failures

The interfaces utilize a flat design language that may obscure interactive elements, leading to weak signifiers and poor affordance. Buttons often resemble static text fields, which can fail to invite interaction. This design choice appears to force operators to rely on trial-and-error exploration or procedural recall to identify and discover control points, rather than recognize intuitively.

Look at this button. It's just an icon with no label. I didn't know it was the insp./exp. hold. No matter how many times I pressed it, the screen didn't show anything to help me figure it out. (Experienced Anesthesia Specialist and ICU Subspecialist).

Theme 2: Cognitive load and Lack of System Transparency

This theme addresses the cognitive demands imposed on the user by the devices. The analysis suggests that the ventilators require high resource utilization from the users, demand on working memory and require mental calculation. This is attributed the disconnect between the system's architecture and the clinician's mental model. This theme explores multiple dimensions of cognitive workload and consists of four subthemes, each of which is described below with illustrative quotes from clinical evaluators.

2.1 Reliance on Error-Prone Memory

The system design appears to diverge from the heuristic principle of minimize memory load. They rely on the users' prospective memory and the ability to remember to perform future actions while automated safeguard is not provided. Some critical tasks, such as terminating nebulization or reverting settings after a maneuver, depend on user's vigilance, which may increase the probability of omission errors under high cognitive load.

We always have to remember to set the time when you're setting the apnea backup. Honestly, I bet most ICU staff forget to do that, so it just ends up staying on the 15-second default. (Experienced ICU Nurse).

2.2 Mental Model Mismatch and Terminology

There appears to be a semantic gap between some devices' logic and clinical reasoning. Participants noted that the terminology, parameter grouping, and iconography often contradicted standard medical conventions. This mismatch requires clinicians to mentally translate their clinical intent into the device's distinctive language, potentially resulting in latency and error in the interaction.

This ventilator's menu is designed differently than what we're used to. You expect to find a parameter in one place, but it has moved elsewhere. For example, I always look for the mode over here, but the layout here just throws me off. I have to literally hunt for everything. It's super confusing. (Experienced anesthesiologist and intensive care specialist).

2.3 Rigid Navigation and High Interaction Cost

Software's architecture is characterized by rigidity and high interaction cost. Simple, high-frequency tasks often require navigation through deep hierarchical structures and mandatory confirmation sequences. This lack of efficiency creates an attention loss, where the user's focus is drawn into navigation tasks rather than maintaining broad oversight of the patient's therapy.

The fact that you have to exit a screen to open another is a real pain. Everything needs an 'Escape' first. Like, if I'm in Alarms and want to check Monitoring, I should be able to just go there directly. See how much of a hassle this is? You're basically stuck in one window until you manually close it. (Master of Science in Critical Care Nursing and staff training supervisor).

2.4 Unclear Logic and Hidden Dependencies

The system frequently restricts user actions through logic that is not immediately apparent to the user. Participants described frustration with parameter constraints where settings are unmodifiable as they are hidden under other variables. The lack of feedback regarding these constraints prevents users from understanding the system's causal relationships, hindering effective troubleshooting.

Sometimes a parameter gets capped, but there's no notification explaining why. It's frustrating that it stays silent..., it's not clear how one change impacts the other parameters... for example in Pediatric mode the Tidal Volume hits a ceiling, and the device won't let me raise the amount, but it doesn't give any reason for the lockout. (Master of Science in Critical Care Nursing and ventilator training supervisor).

Theme 3: Ineffective Error Prevention mechanism and Absence of Safety Constraints

This theme characterizes the ventilators as an under-constrained environment regarding safety resilience. Some devices appear to lack elements of a collaborative safeguard, permitting hazardous inputs without challenge. The findings highlight a deficiency in safety interlocks and reliable automation, creating an environment where latent errors may accelerate into active failures. This theme includes three subthemes, which are presented below along with representative quotes that further clarify their meaning.

3.1 defect of Error-Prevention Mechanisms

A primary safety concern identified is the lack of hard constraints or interlocks for critical parameters. The device allows the entry of settings that may be physiologically critical (e.g., zero respiratory rate for alarm limit, and disabling vital sensors (flow and O2 sensor)) without adequate confirmation barriers. This open interaction logic transfers the entire responsibility of vigilance to the human operator, which may not fully account for human performance limitations in high-stress environments.

The problem is that this device lets you set everything from zero to the max. Why am I allowed even set the RR to zero? That's like giving the patient no breath at all. What if I mess up due to fatigue of night and long shifts? The device should prevent me from making errors and not allow zero as an option. It needs safety min and max, but unfortunately, it just gives you the full range with no restrictions. (Master of Science in Critical Care Nursing).

3.2 Automation Discrepancies and State Transitions

Some devices' automation logic, especially regarding state transitions like standby, start-up, and apnea backup, was reported as inconsistent. This led to unanticipated system behaviors. Participants reported instances where the device, in case the operator forgets, failed to trigger ventilation upon patient connection or failed to revert to the primary mode post-apnea. These discrepancies disrupt the operational state, leaving the clinician incorrectly under the assumption the patient is being ventilated safely.

The catch with this ventilator is that it stays in apnea backup even after the patient starts breathing again. You have to manually hit 'Alarm Reset' to get it back to SPONT mode. This is a huge issue for me. I warning the nurses for nearly a decade, and yet it still happens. I happened to see patient stuck in backup mode for hours just because the reset button wasn't pressed. (Experienced internist and pulmonary specialist and 9 of other participants).

3.3 Deficiencies in Alarm Logic and Error Recovery

The alarm system often fails to support effective error recovery. Alarms use technical codes rather than clinical descriptions. Besides, as alarm are not descriptive enough the users fail to take needed which contributing to alarm fatigue and desensitization. Furthermore, the absence of onboard decision support leaves clinicians without guidance during critical failure modes, potentially increasing the time to resolution.

The alarm shouldn't just keep flashing 'High Pressure, High Pressure.' We know the pressure is high, it's an ARDS case! It needs to provide some troubleshooting tips instead. It should be able to tell if the tube is kinked and prompt me to 'Check the circuit' or 'Check the patient.' It would be so much better if it listed the top reasons for the alarm so I could tackle them one by one instead of spending a lot of time trying to find the cause. (experienced ICU nurse).

Theme 4: Operational Inefficiency and Contextual Incompatibility

This theme underscores the importance of macro ergonomics and the context of use. The data suggests some device may have been designed without adequate consideration for the sociotechnical systems of the ICU. It conflicts with established clinical workflows. It presents physical hazards within the crowded workspace, creating operational impediments that necessitate workarounds. This theme comprises two subthemes, which are presented below together with illustrative quotes from the evaluators.

4.1 Workflow Inefficiency and Clinical Inflexibility

Some devices exhibit a lack of clinical flexibility, enforcing rigid workflows that do not align with diverse patient needs (e.g., pediatric vs. adult variations) or standard protocols (e.g., weaning procedures). This inflexibility forces clinicians to adapt their practice to the tool, rather than the tool supporting the practice, resulting in inefficient care delivery.

The 'pre-use check' setup is such a drag; it takes so long. And is ridiculous. It consist of several pages long, so most users skip them. Hardly anyone actually does the whole thing properly. If a feature is so long-winded that everyone skips, it's just not practical. In high-stress moments when we have an emergency admission and need to get a patient on a ventilation fast, these long-drawn-out steps are the last thing you want to deal with. (Experienced ICU nurse).

4.2 Physical Features and Maintainability Challenges

The hardware design presents significant physical interface mismatches. Issues such as bulky structure, side-mounted connectors, and poor breath circuits management contribute to workspace obstruction and risking accidental extubation. Additionally, the fragility of components and complex maintenance requirements impose a significant effort on staff, impacting the device's availability and long-term reliability.

The way the tubes are positioned is a real problem in the ICU. They get stretched, forcing patients to stay in one spot for way too long. Since the device can only go on one side of the bed, the ETT often gets tense or even dislodged. The tubing also gets in the way when nurses try to perform suctioning, making it less effective. If a patient moves, the ETT is stuck. Then it will be at the point where the equipment is actually interfering with the treatment. (Experienced ICU nurse and Supervisor of Blood Gas department and 7 of evaluators).

4. Discussion

The primary objective of this study was to assess the usability of seven ICU ventilators through a heuristic evaluation. A total 42 in-situ interview sessions conducted across ten university-affiliated educational hospitals, ensuring assessment within real clinical environments. Experienced clinicians participated as heuristic evaluators, representing a predominantly senior workforce with a mean age of 53.5 years and over 13 years of ventilator work experience.

Analysis showed that the most frequent heuristic violations involved Match, Error, and Visibility. The prominence of ‘Match’ differs from prior studies that identified Consistency as the dominant issue [16, 25], likely reflecting clinicians’ heightened sensitivity to alignment between device terminology and clinical workflows rather than purely graphical inconsistencies. This finding supports recommendations that interface logic should correspond to users’ mental models [1]. Recurrent violations in Error and Visibility further indicate deficits in system transparency rather than mere layout flaws, effectively turning ventilators into unclear black boxes that conceal critical safety states [13]. Such deficiencies likely contribute to the high task-failure rates reported in related studies of ventilator usability [26].

Risk profiles varied markedly across ventilators, indicating that usability functions as a key design differentiator rather than an industry baseline. More functionally complex devices showed higher problem frequencies, supporting the view that poor usability often arises from unmanaged complexity and engineering-driven design priorities [26]. The presence of severe defects in commercially approved systems further shows that regulatory compliance alone does not ensure good usability [18]. In this regard, expert heuristic inspection was particularly effective in uncovering critical use errors that are often missed in brief simulated-use testing or concealed by experienced users’ adaptive behaviors [16].

Experience analysis of evaluators revealed a clear pattern in usability assessment. Although mean team experience did not correlate with the total number of problems identified, higher-tenured clinicians were significantly less likely to rate issues as major or catastrophic, likely reflecting cognitive desensitization, where experts internalize workarounds and compensate for design flaws [1]. These findings caution that exclusive reliance on experienced users may underestimate risks for less experienced staff, as accumulated expertise can mask critical design deficiencies and inflate usability ratings, highlighting a potential limitation in participatory design approaches [10, 27, 28].

The physician-to-nurse ratio, the proportion of physicians in the teams, and their correlated with the frequency or severity of identified usability problems indicate that fundamental interaction challenges affect all users similarly, regardless of clinical role. While differing from prior studies that focused on homogenous groups [29] or role-based performance differences [13], these findings align with Wachter et al. [30], confirming that systemic interface flaws impose a consistent burden across professional hierarchies.

Inter-rater variability in severity ratings indicates that probably perceptions of usability defects are shaped by individual mental models, clinical experience, and risk tolerance. While prior studies using homogenous raters or binary metrics reported higher reliability [16, 31], this study’s multidisciplinary team and a five-point scale captured the real-world diversity of risk perception in clinical settings. These findings underscore that single-evaluator judgments may bias results and highlight the importance of aggregating multiple expert opinions or consensus-based scoring to improve evaluation validity [18, 20, 32].

The findings of thematic analysis indicate that the usability deficiencies in the ICU ventilators are not merely cosmetic but structural. They stem from a failure to account for the perceptual limits, cognitive

biases, and environmental constraints of the human operator. The theme one highlights excessive perceptual demands imposed by certain device interfaces. Key deficiencies included poor legibility and low contrast, hazards exacerbated by the dim lighting common in ICUs. Additionally, the use of flat design visual languages often obscured interactive signifiers, reducing critical affordances [11, 14]. Beyond visual ergonomics, nested menus frequently concealed system states, such as battery life or sensor status, inducing operator uncertainty. These findings indicate that prioritizing aesthetic simplicity over comprehensive feedback compromises situational awareness and diminishes overall usability [13].

The second theme, highlights how ventilator interfaces impose high demands on operators' working memory. Key contributors included reliance on prospective memory for critical tasks without automated safeguards, rigid navigation through deep menus, and non-standardized terminology and icons, all of which require real-time mental translation and elevate the risk of use errors [14, 29]. While pattern-based graphical displays can aid interpretation [33], oversimplifying interfaces or using unvalidated touchscreen designs compromises clinical safety [34, 35]. Devices with redundant or non-essential features further increased cognitive load and received lower usability ratings, indicating that feature creep negatively biases operator performance [36].

The third theme, highlights critical safety vulnerabilities in ventilator design. Several devices lacked forcing functions or data validation, allowing physiologically hazardous settings without objection [8]. Automation ambiguities, such as failure to trigger ventilation upon patient connection, combined with alarms that signal malfunctions without actionable guidance, leave clinicians unsupported. As a result, operators must rely entirely on their own troubleshooting during high-pressure events, increasing the risk of active failures [10, 12].

The fourth theme emphasizes the importance of macro-ergonomics in ventilator design. Devices are often developed in isolation, misaligned with ICU workflows, forcing users to create workarounds such as disabling essential sensors [37]. Physical interface challenges, including poor tubing management, bulky form structure, and complex maintenance, further increase workload and can compromise operational viability, particularly in resource-limited settings [38, 39]. The four thematic areas can help manufacturers have a roadmap for improvement. The goal is to evolve the ventilator from a complex machine that requires constant management into a transparent, resilient clinical partner that enhances, rather than impedes, the delivery of critical care.

The results indicate that deploying practicing physicians and nurses as evaluators within their actual work environment is an effective strategy for uncovering latent usability flaws and predicting patient safety risks. Integrating end-user's clinical expertise with established heuristic principles via semi-structured interviews and device walkthroughs enabled the identification of high-severity issues, including missing functionalities and workflow inconsistencies, beyond superficial interface anomalies [16]. Unlike studies relying on lay users or simple error counts [13, 40], this approach provided a granular, accurate classification of usability problems [38]. By training ventilator professionals in heuristic

methods, we created dual-domain experts, bridging the gap between generalist evaluators and specialized medical devices, and yielding more reliable insights into workflow disruptions [20, 31].

Study Limitations

Several limitations should be considered when interpreting these findings. First, the brief training provided to clinical evaluators on heuristic principles may have influenced the depth and precision of their assessments. Second, variability in severity ratings reflects the inherent subjectivity of expert inspection, influenced by individual mental models and risk tolerance, highlighting the need to triangulate heuristic results with objective performance metrics. Third, the number of evaluators per device varied (highest for device R, lowest for device DEX), which may have affected the total problems identified; this variation arose from challenges in recruiting clinicians both highly experienced with specific ventilators and willing to participate. Fourth, some catastrophic usability issues were detected in certain devices but overlooked or rated lower in others, likely due to fluctuations in evaluator attention. Finally, four of the ventilators, although widely used in hospitals, were not the newest models from their manufacturers, who have since released updated versions.

Implications for Industry and Design

The empirical insights derived from this study offer critical strategic guidance for the engineering and industrial design of mechanical ventilators. Our analysis demonstrates that strict adherence to technical standards and regulatory compliance is insufficient to guarantee operational safety within high-complexity clinical environments. Consequently, it is recommended that manufacturers integrate the following Human Factors Engineering principles into their product development.

1. Design priority need to shift from superficial visual consistency to the alignment of the device's logic with the clinician's established mental models. This necessitates an industry-wide commitment to the harmonization of terminology, iconography, and units of measurement. By standardizing these semantic elements, manufacturers can significantly reduce the cognitive load associated with the real-time mental translation of device data.
2. In accordance with the hierarchy of hazard controls, safety should be engineered directly into the system architecture rather than delegated to administrative controls such as textual warnings or user training. Manufacturers should implement intelligent forcing functions and safety interlocks to preclude hazardous physiological inputs, such as the inadvertent deactivation of vital sensors. Furthermore, critical automation states, specifically Standby and Apnea Backup modes, require explicit, unambiguous feedback or warning mechanisms to ensure accurate mode awareness.
3. Validation activities must extend beyond controlled laboratory settings to encompass the rigorous physical and environmental constraints of the actual ICU context of use. Design teams need to take these factors into account: low ambient lighting, spatial restrictions, acoustic interference, and the performance decrements associated with operator fatigue and circadian disruption. Regarding the user interface, the use of high-contrast typography is essential for legibility, while flat design elements that obscure affordances should be avoided. Hardware architecture such as filter

exchange, must also facilitate maintenance tasks without inducing physical strain or workflow disruption.

4. User engagement processes should recruit a diverse range of participants across the experience continuum, from novice residents to senior specialists. As our findings indicate, highly proficient users often develop adaptive strategies that mask systemic defects. Therefore, soliciting feedback from less experienced clinicians is essential to identify latent risks that may remain invisible to subject matter experts.

Suggestions for Future Studies

Future research should advance the field by triangulating subjective heuristic data with objective psychophysiological metrics (e.g., eye-tracking) and employing longitudinal designs to examine the mechanisms of expert desensitization. Methodological precision requires the development of domain-specific heuristics tailored to critical care, alongside structured consensus frameworks (e.g., Delphi) to mitigate inter-rater variability. Finally, adopting a systemic macro-ergonomic perspective is crucial; studies should be replicated across diverse healthcare infrastructures and include the broader stakeholders to ensure that proposed design remediations are operationally feasible within the complex socio-technical context of the ICU.

5. Conclusions

The findings of this study demonstrate that ventilator use errors primarily arise from fundamental design deficiencies rather than operator incompetence. Even highly experienced clinicians face substantial interaction barriers, highlighting that training alone cannot resolve the misalignment between the user's mental model and the device's operational logic. This underscores the necessity of designing ventilators that reduce cognitive load, harmonize interface terminology, and provide clear, actionable feedback. Additionally, usability evaluations should extend beyond controlled settings to real-world clinical environments to uncover latent hazards. Ultimately, ensuring ventilator usability is not optional but a critical prerequisite for patient safety, supporting the development of devices that are both effective and resilient in high-pressure ICU settings.

Declarations

Acknowledgments

This study was conducted as part of the requirements for the PhD degree in Ergonomics at Shiraz University of Medical sciences. The authors sincerely thank the following clinicians for their invaluable contribution to this study: Dr. Ghasemzadeh, Dr. Alinejad, Dr. Zand, Dr. Sabetian, Dr. Haddad Bakhodaei and ICU Nurses Mr. Bonyani Podonak, Ms. Izadi, Mr. Razmjooie, Mr. Hosseini, Mr. Abedi, Ms. Elahi, Mr Rezvan, Mr. Sadeghi Mehr, Ms. Karamad, Ms. Rasekhi, Mr. Ghiyasi, Ms. Niakan, Mr. Rajabi, Ms. Afsharnia, Mr. Yazdani, Mr. Ahmadloo. Their expertise and time in conducting the heuristic evaluations were essential for identifying critical usability issues and informing our findings.

Author contributions

S. Ahmadi Charkhabi: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. **H. Parsaei:** Conceptualization, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing. **S. M. Masoompour:** Conceptualization, Methodology, Project administration, Resources, Validation, Writing – review & editing. **H. Azizzadeh:** Conceptualization, Methodology, Project administration, Validation, Writing – review & editing. **J. Hassanzadeh:** Conceptualization, Formal analysis, Methodology, Validation, Writing – review & editing. **A. Massoudi Rad:** Investigation, Resources, Validation, Writing – review & editing. **S. M. Mortazavi:** Conceptualization, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **A. Choobineh:** Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Validation, Writing – review & editing.

Funding

This study was financially supported by Shiraz University of Medical Sciences via contract number 28681. The funding body had no role in the design of the study, data collection, analysis, interpretation of data, or in writing the manuscript.

Additional information

Disclosure Statement

H.A. is employed by a medical device company that manufactures one of the ventilator models evaluated in this study. To mitigate potential bias, this author was not involved in data collection, usability evaluations, severity rating, data analysis, or interpretation of the results related to the evaluated devices. All evaluations and analyses were conducted independently by the remaining authors. The authors confirm that this employment relationship did not influence the study design, data collection, analysis, or reporting of the findings.

Correspondence and requests for materials should be addressed to A.Choobineh.

Reprints and permissions information is available at www.nature.com/reprints

References

1. Hignett, S. et al. Human factors/ergonomics to support the design and testing of rapidly manufactured ventilators in the UK during the COVID-19 pandemic. *Int. J. Qual. Health Care*. **33**, 4–10. https://doi.org/10.1093/intqhc/mzaa089 (2021).
2. Iyengar, K., Bahl, S., Vaishya, R. & Vaish, A. Challenges and solutions in meeting up the urgent requirement of ventilators for COVID-19 patients. *Diabetes Metabolic Syndrome: Clin. Res. Reviews*. **14**, 499–501. https://doi.org/10.1016/j.dsx.2020.04.048 (2020).

3. Xie, M., Liu, X., Cao, X., Guo, M. & Li, X. Trends in prevalence and incidence of chronic respiratory diseases from 1990 to 2017. *Respiratory research* 21, 49, (2020). https://doi.org/10.1186/s12931-020-1291-8
4. Viegi, G., Maio, S., Fasola, S. & Baldacci, S. Global burden of chronic respiratory diseases. *J. aerosol Med. pulmonary drug delivery*. **33**, 171–177. https://doi.org/10.1089/jamp.2019.1576 (2020).
5. Cassidy, C., Smith, A. & Arnot-Smith, J. Critical incident reports concerning anaesthetic equipment: analysis of the UK National Reporting and Learning System (NRLS) data from 2006–2008. *Anaesthesia* **66**, 879–888 (2011).
6. Kamio, T. & Masamune, K. Mechanical ventilation-related safety incidents in general care wards and ICU settings. *Respir. Care*. **63**, 1246–1252. https://doi.org/10.4187/respcare.06109 (2018).
7. Vignaux, L., Tassaux, D. & Jolliet, P. Evaluation of the user-friendliness of seven new generation intensive care ventilators. *Intensive Care Med.* **35**, 1687–1691. https://doi.org/10.1007/s00134-009-1580-7 (2009).
8. Uzawa, Y., Yamada, Y. & Suzukawa, M. Evaluation of the user interface simplicity in the modern generation of mechanical ventilators. *Respir. Care*. **53**, 329–337. https://doi.org/10.4187/respcare.08530329 (2008).
9. Kacmarek, R. M. The mechanical ventilator: past, present, and future. *Respir. Care*. **56**, 1170–1180. https://doi.org/10.4187/respcare.01420 (2011).
10. Spaeth, J., Schweizer, T., Schmutz, A., Buerkle, H. & Schumann, S. Comparative usability of modern anaesthesia ventilators: a human factors study. *BJA: Br. J. Anaesth.* **119**, 1000–1008. https://doi.org/10.1093/bja/aex226 (2017).
11. Jiang, M., Sun, D., Li, Q. & Wang, D. The usability of ventilator maintenance user interface: A comparative evaluation of user task performance, workload, and user experience. *Sci. Prog.* **103**, 0036850420962885. https://doi.org/10.1177/0036850420962885 (2020).
12. Zheng, Q., Liu, X. & Wang, W. A consensus model-based risk matrix for human error factors risk analysis in medical devices by considering risk acceptability. *Reliab. Eng. Syst. Saf.* **238**, 109446. https://doi.org/10.1016/j.ress.2023.109446 (2023).
13. L'Her, E. & Roy, A. Bench tests of simple, handy ventilators for pandemics: performance, autonomy, and ergonomy. *Respir. Care*. **56**, 751–760. https://doi.org/10.4187/respcare.00841 (2011).
14. Coldewey, B., Diruf, A., Roehrig, R. & Lipprandt, M. Causes of use errors in ventilation devices-Systematic review. *Appl. Ergon.* **98**, 103544. https://doi.org/10.1016/j.apergo.2021.103544 (2022).
15. Coombs, E. The usability of ventilators: a comparative evaluation of use safety and user experience—an editorial response. *Crit. Care*. **21**, 164. https://doi.org/10.1186/s13054-017-1751-9 (2017).
16. Zhang, J., Johnson, T. R., Patel, V. L., Paige, D. L. & Kubose, T. Using usability heuristics to evaluate patient safety of medical devices. *J. Biomed. Inform.* **36**, 23–30. https://doi.org/10.1016/S1532-0464(03)00060-1 (2003).
17. Templier, F. et al. Evaluation of the ventilator-user interface of 2 new advanced compact transport ventilators. *Respir. Care*. **52**, 1701–1709. https://doi.org/10.4187/respcare.07521701 (2007).

18. Martin, J. L., Norris, B. J., Murphy, E. & Crowe, J. A. Medical device development: The challenge for ergonomics. *Appl. Ergon.* **39**, 271–283. https://doi.org/10.1016/j.apergo.2007.10.002 (2008).
19. Flewwelling, C. J., Easty, A. C., Vicente, K. J. & Cafazzo, J. A. The use of fault reporting of medical equipment to identify latent design flaws. *J. Biomed. Inform.* **51**, 80–85. https://doi.org/10.1016/j.jbi.2014.04.009 (2014).
20. Klarich, A. et al. Usability of smart infusion pumps: A heuristic evaluation. *Appl. Ergon.* **98**, 103584. https://doi.org/10.1016/j.apergo.2021.103584 (2022).
21. Food & Administration, D. Applying human factors and usability engineering to medical devices: guidance for industry and Food and Drug Administration staff. *Food and Drug Administration: Rockville, MD, USA* (2016).
22. Sujan, M. et al. The contribution of human factors and ergonomics to the design and delivery of safe future healthcare. *Future Healthc. J.* **8**, e574–e579. https://doi.org/10.7861/fhj.2021-0112 (2021).
23. Clarke, V. & Braun, V. Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *The psychologist* 26 (2013).
24. Ahmed, S. K. et al. Using thematic analysis in qualitative research. *J. Med. Surg. Public. Health*. **6**, 100198. https://doi.org/10.1016/j.glmedi.2025.100198 (2025).
25. Silva, R., Lima, A. C., Andrade, E., Martins, A. I. & Rocha, N. P. Heuristic evaluation of the usability of a mechanical ventilator interface through a simulator. *Procedia Comput. Sci.* **219**, 1232–1239. https://doi.org/10.1016/j.procs.2023.01.406 (2023).
26. Marjanovic, N. S., De Simone, A., Jegou, G. & L'her, E. A new global and comprehensive model for ICU ventilator performances evaluation. *Ann. Intensiv. Care.* **7**, 68. https://doi.org/10.1186/s13613-017-0285-2 (2017).
27. Buehler, P. K. et al. Differing Visual Behavior Between Inexperienced and Experienced Critical Care Nurses While Using a Closed-Loop Ventilation System—A Prospective Observational Study. *Front. Med.* **8**, 681321. https://doi.org/10.3389/fmed.2021.681321 (2021).
28. Calvano, M., Curci, A., CM, N. F., Piccinno, A. & Sohail, Q. in *Proceedings of the 1st International Workshop on Participatory Design & End-User Development-Building Bridges (PDEUD*). (2024).
29. Jiang, M., Liu, S., Feng, Q., Gao, J. & Zhang, Q. Usability study of the user-interface of intensive care ventilators based on user test and eye-tracking signals. *Med. Sci. monitor: Int. Med. J. experimental Clin. Res.* **24**, 6617. https://doi.org/10.12659/MSM.909933 (2018).
30. Wachter, S. B. et al. The evaluation of a pulmonary display to detect adverse respiratory events using high resolution human simulator. *J. Am. Med. Inform. Assoc.* **13**, 635–642. https://doi.org/10.1197/jamia.M2123 (2006).
31. Katre, D., Bhutkar, G. & Karmarkar, S. in *IFIP Working Conference on Human Work Interaction Design*. 83–97 (Springer).
32. Colman, N. et al. Simulation-based user-centered design: an approach to device development during COVID-19. *Pediatr. Qual. Saf.* **6**, e427. https://doi.org/10.1097/pq9.0000000000000427 (2021).

33. Liu, Y., Tech, L. & Osvalder, A. L. Usability evaluation of a GUI prototype for a ventilator machine. *J. Clin. Monit. Comput.* **18**, 365–372. https://doi.org/10.1007/s10877-005-7997-9 (2004).
34. Lin, M. C. M. et al. Involvement of frontline clinicians in healthcare technology development: Lessons learned from a ventilator project. *Health Technol.* **12**, 597–606. https://doi.org/10.1007/s12553-022-00655-w (2022).
35. Gaben, S. S. M. et al. ExoventQ: A novel low-cost portable negative pressure ventilator design and implementation. *IEEE Access.* **12**, 61828–61847. https://doi.org/10.1109/ACCESS.2024.3390566 (2024).
36. Black, C. L., Morris, A. E., Akinluyi, E. & Kay, R. H. Heuristic evaluation of portable pulse oximeters for domiciliary use: Implications for its use in assessing medical device usability. *Smart Health*. **27**, 100357. https://doi.org/10.1016/j.smhl.2022.100357 (2023).
37. Furniss, D., Masci, P., Curzon, P., Mayer, A. & Blandford, A. 7 Themes for guiding situated ergonomic assessments of medical devices: A case study of an inpatient glucometer. *Appl. Ergon.* **45**, 1668–1677. https://doi.org/10.1016/j.apergo.2014.05.012 (2014).
38. Morita, P. P. et al. The usability of ventilators: a comparative evaluation of use safety and user experience. *Crit. Care*. **20**, 263. https://doi.org/10.1186/s13054-016-1431-1 (2016).
39. Tayler, A. et al. Feasibility of a novel ultra-low-cost bubble CPAP (bCPAP) System for neonatal respiratory support at Muhimbili National Hospital, Tanzania. *Plos one*. **17**, e0269147. https://doi.org/10.1371/journal.pone.0269147 (2022).
40. Marjanovic, N. & L'Her, E. A comprehensive approach for the ergonomic evaluation of 13 emergency and transport ventilators. *Respir. Care*. **61**, 632–639. https://doi.org/10.4187/respcare.04292 (2016).

Figures

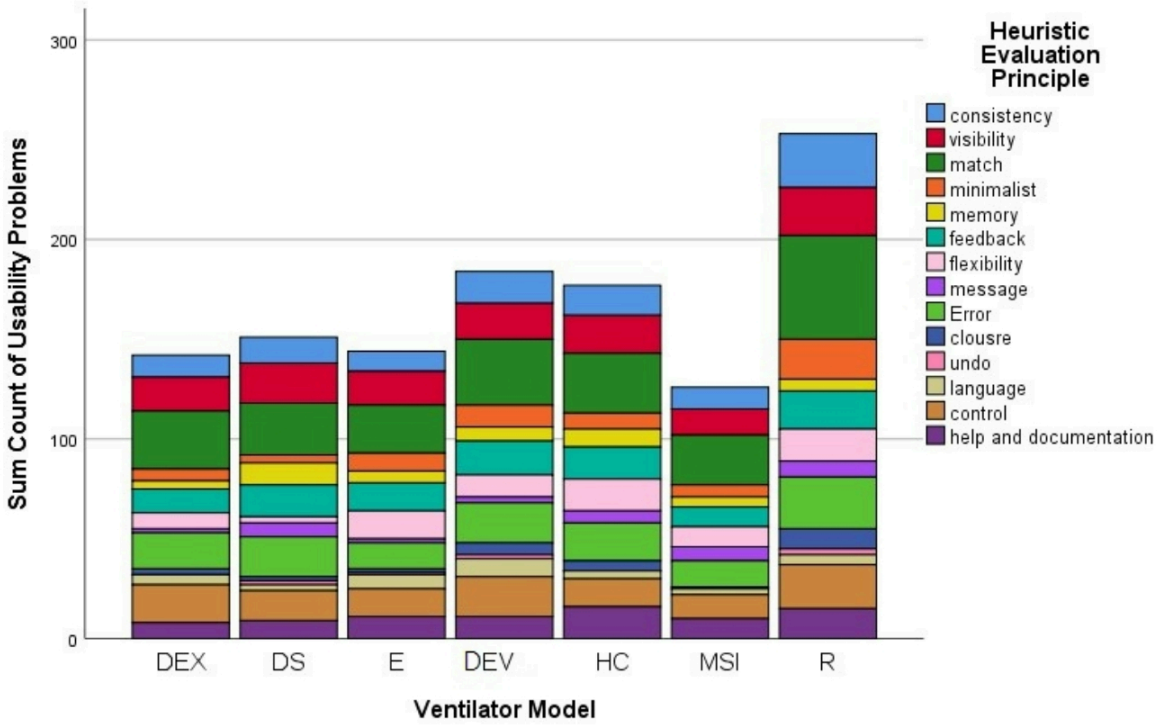


Figure 1

Total Count of Usability Problems by Ventilator Model and Heuristic Principle. The height of each bar represents the total number of identified usability problems for each ventilator. The colored segments show the contribution of each of the 14 heuristic evaluation principles to the total count.

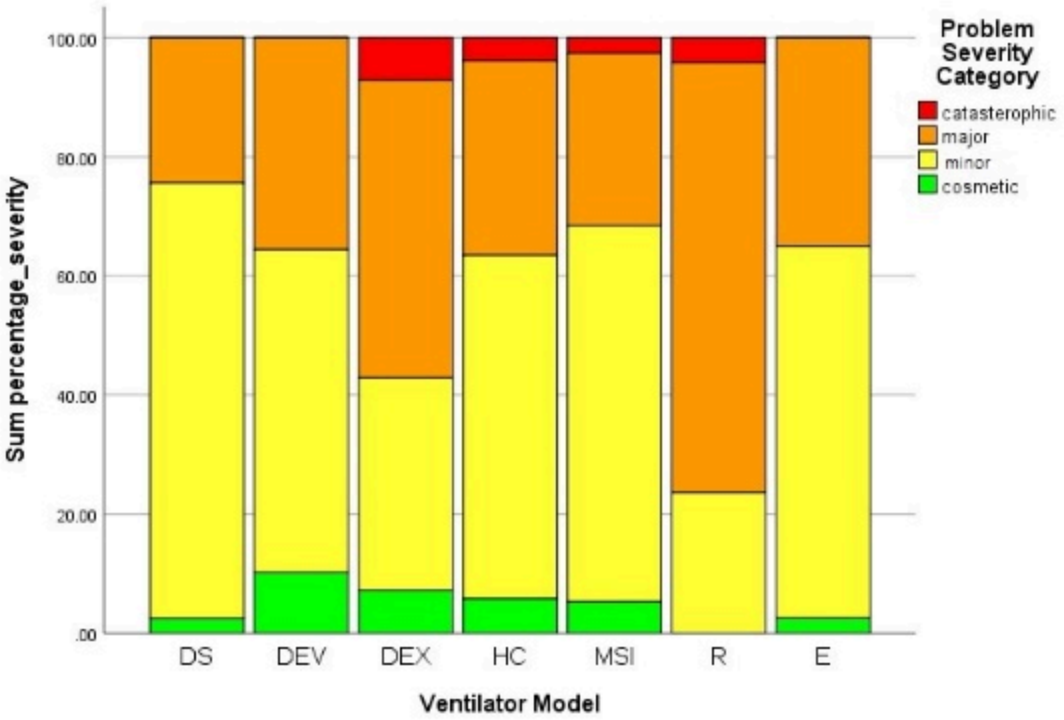


Figure 2

Comparison of usability problem severity profiles across the seven ventilator models. Stacked bar chart showing the percentage distribution of usability problem severity (cosmetic, minor, major, catastrophic) across seven ICU ventilator models.