MEDICAL SCIENCE MONITOR

CLINICAL RESEARCH

e-ISSN 1643-3750
© Med Sci Monit, 2018; 24: 6617-6629
DOI: 10.12659/MSM.909933

Received: 2018.03.13
Accepted: 2018.06.05
Published: 2018.09.20

Usability Study of the User-Interface of Intensive Care Ventilators Based on User Test and Eye-Tracking Signals

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

ACDE 1,2 **Mingyin Jiang**
AFG 1,2 **Shenglin Liu**
BC 1,2 **Qingmin Feng**
BCD 1,2 **Jiaqi Gao**
ADEFG 1,2 **Qiang Zhang**

1 Department of Medical Engineering, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, P.R. China
2 Healthcare Ergonomics Lab, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, P.R. China

Corresponding Author: Qiang Zhang, e-mail: whuhcce@163.com
Source of support: This study was supported by the National Key R&D Program of China (Number: 2016YFC0106702)

Background: The aim of this study was to evaluate the ergonomics of the user-interface for 3 intensive care ventilators, and identify usability problems leading to user errors.

Material/Methods: Sixteen respiratory therapists were recruited to perform 6 specific user-interface operational tasks on ventilators. Data (task completion time, pupil diameter, average slope of pupil diameter change, and subjective evaluation) were collected through objective measurement, questionnaires, and an eye-tracking instrument.

Results: For task completion time, there were significant differences among ventilators in recognition tasks of ventilator mode and settings ($P<0.05$), modification of ventilator modes and recognizing ($P<0.05$) and changing alarm settings ($P<0.05$). A mean of 15±2 task failures was observed for each ventilator. For the change of pupil diameter, a significant difference was observed between ventilators (except task 2, $P<0.05$). For average slope of pupil diameter change, a significant difference was also observed between ventilators (except task 2, $P<0.05$). The Servo I showed a better correlation between task completion time and pupil diameter change. The subjective evaluation results were clear: Evital 4 received worst scores in terms of friendliness of user-interface, information display and safety (respectively, $P<0.05$).

Conclusions: The present study provided valuable evidence to indicate the ergonomic of ventilators now used in China. With the result of this study, we can infer that the Evital 4 were poorly ergonomic designed. Furthermore, the study also demonstrated that eye-tracking can be a promising tool to evaluate the ergonomics of the user-interface.

MeSH Keywords: **Critical Care • Equipment Design • Equipment Safety • Eye Movements • User-Computer Interface • Ventilators, Mechanical**

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/909933

3690 11 1 40

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Background

Hospitals often rely on ventilators in intensive care units (ICUs) to support respiration in fragile patients. However, ventilators are complex devices and small errors in adjustment of ventilation parameters can lead to significant morbidity. A significant risk with ventilators is use-related error [1–3]. According to published studies, ventilator use-related errors had a high frequency of occurrence [4]. For healthcare providers in ICUs, interaction with ventilators occupied approximately 25% of the daily work [5] and was one of the major activities leading to use error [5,6]. Use error could result in patient injury or even death if machines have no appropriate design to eliminate these risks [7,8]. It is important, therefore, to encourage the use of ventilators that have been designed to eliminate these risks.

It is common to direct blame for adverse events toward the healthcare provider, but machine design defects have been shown to contribute significantly to use-related errors, and user-interface is one of the root causes [9–11]. Healthcare providers operate medical devices through user-interfaces that have powerful ability to reduce and eliminate adverse events [12–14]. The evaluation of the user-interface can be carried out by usability testing [15,16]. Usability testing has been employed in medical devices design in recent years, including by the US Food and Drug Administration, who developed a guidance document to assist manufacturers improving medical device design of the user-interface to minimize use error at the regulatory level [15]. To address use errors at the user-interface, several usability studies of ventilators have been conducted, and results proved that poor ergonomic design of the user-interface resulted in operational delays and user errors [17–22]. However, the test tasks were always time-urgent and were sequential operations, making it difficult to evaluate or score in a timely fashion. Moreover, these studies evaluated ergonomics of the user-interface mainly based on task completion time, failures, and mental workload of the participant. These parameters heavily rely on experts' or participants' ability. Therefore, with a gap in experts' ability to objectively evaluate the performance of participants during test, it is difficult to acquire reliable and credible results for user-interface assessment of ventilators.

Based on eye movement features, eye tracking may be a possible solution to these problems. Eye movement features can be used as an adjunct to standardize evaluation methods to overcome the difficulties described above. Eye tracking has already been used in medical fields regarding the usability of modern anesthesia ventilators evaluation [23], in the workload of laparoscopic surgeons [24], in nursing education [25], and in the ergonomic evaluation of ventilator [26]. Particularly for pupil dilatation, the technique has been shown to be sensitive to task difficulty and workload [27,28]. Published studies have shown that larger pupil sizes change from baseline implies greater task difficulty and mental workload [24,26,29–31]. Therefore, the measurement of pupil diameter can be used as a standardized evaluation method for assessing the ergonomic of user-interface of ventilators.

We performed complete usability testing of 3 intensive care ventilators to evaluate their ergonomics, taking into account pupil diameter as a standardized assessment indicator.

Material and Methods

Study design

This was a prospective crossover usability study of respiratory therapist performing 6 typical user-interface operation tasks on 3 ventilators in a simulated clinical usage environment in a hospital ICU in Wuhan, Hubei Province, China. The objective of this study was to evaluate the user-interfaces of the machines, rather than to assess the participants' performance. The procedures in this study were performed to follow the local treatment protocols. This study has been approved by the Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology (IORG No: IORG0003571).

Participants

Sixteen respiratory therapists were recruited from the ICU of our hospital. Before the formal test, we were provided the tested ventilator (Evita 4, Servo I, and Boaray 5000D) operation training for all participants. Participants were given a series of learning goals such as settings value modification, ventilation mode change, alarm settings value modification, and menu browse. A tester was available to participants to explain the function of the ventilator and to answer questions. Informed consent was obtained from all participants prior to the study. According to the published studies, medical device usability studies that involved 8 to 12 participants can obtain reliable findings [15,32,33]. Before the formal test, a preliminary study with 4 participants was performed to verify and improve the test flow and to analyze the reliability of test collection data.

Ventilators

Three ICU ventilators, which are commonly used in the ICU department of our local medical institutions, were selected and made available for our usability study: Evita 4 (Draeger, Lubeck, Germany; software version: 04.24 07/12/11), Servo I (Maquet, Solna, Sweden; software version: v5.00.00), and Boaray 5000D (Probe, Shenzhen, China; software version: 0A_006_V06.10.02_151119). New generation ventilators, such as the

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

V500 (Draeger) and Servo U (Maquet), have been developed by the manufacturers. However, these new generation ventilators are rarely used in our local medical institutions, making them unavailable for our test. The Boaray 5000D is a similar product as Evita 4 and Servo I, which has been used in nearly a thousand hospitals in China and has been exported to more than 50 foreign countries and regions. Therefore, the test ventilators in our study represent current ventilator in our local region, and available for our usability study. In the study, each ventilator was equipped with a standard double limb circuit and connected to a test lung (Venti.Plus™, GaleMed, Taipei, Taiwan, China).

Test tasks

A total of 6 typical user-interface operation tasks on the ventilator were considered: 1) recognition of ventilator mode and settings (reading of the ventilator mode and setting values set by the tester); 2) recognition of monitored values (minute volume, respiratory rate, positive end-expiratory pressure, and tidal volume); 3) modification of ventilator settings (FIO_2, VT, RR, PEEP, and P_{insp}); 4) modification of ventilator modes (from VC-IMV to PC-CSV or from PC-CSV to VC-IMV); 5) recognizing and changing alarm settings (minute volume, respiratory rate, airway pressure); and 6) respond to alarm (identify the reason for the alarm and manage it). For more details, please see supplementary materials. A detailed list for each test task was developed to ensure the consistency of test processes for each ventilator.

For each test task, the participants were allowed only one attempt, and an upper time limit for task completion was set; participants needed to give the correct response within 180 seconds [18,20,22]. A test task was identified as a failure if participants did not make the right response, if their response exceeded upper the time limit, or if they abandoned the task.

Eye-tracking data recording and analyses

We sampled pupil diameter at 50 Hz using the Tobii Glasses 2 Eye Tracker (Tobii Technology, Danderyd, Sweden) as the eye-tracking instrument. Before the participants performed each test task on the ventilator, they were instructed to perform a pupil calibration process according to manufacturer recommendations. At this time, we measured baseline pupil diameter (initial diameter) for each participant. We adjusted the raw pupil diameter data according to the participant's baseline pupil diameter. The change pupil diameter from baseline, as a result of participant's mean pupil diameter change, acted as an objective indicator to evaluate the ergonomic of user-interface among different ventilators [26]. Furthermore, the average slope of pupil diameter change from baseline over time (for each task) [34] was also calculated and matched with pupil diameter changes to evaluate the ergonomics of the ventilators' user-interface.

Subjective evaluation

The 3 subjective evaluation points for the ventilator user-interface were assessed by participants via questionnaire, using questions regarding user-friendliness of user-interface (How user-friendliness was the user-interface of the machine?) information display friendliness of user-interface (How do you evaluate the information display of the user-interface of the machine?); and safety of the user-interface (How do you evaluate the safety of the user-interface of the machine?). The answers to these questions were ranked using Likert scales, with a final score ranging from 1 (worst value) to 5 (best value).

Study procedure

This study was conducted in an ICU treatment room, at a tertiary hospital in Hubei Province of central China. Participants were requested to accomplish the test tasks in test devices following local treatment protocols.

The test ventilators were assigned randomly (Supplementary Table 1). Participants needed to be equipped with the eye tracker and to perform a calibration process prior each task starting. The diameter of the pupil is affected by many factors, including illumination and bodily movement. To ensure accurate determination of pupil diameter, we took several measures, including maintaining the intensity of light at a constant level; participants were required to minimize bodily movements and to avoid changes in direction of gaze. Once the study started, a tester standing near the participant would inform the participant of the test task. The participant made only one attempt to perform the test task. Once the task was completed, the task completion time and eye-tracking data were recorded. Simultaneously, an expert would check completion status of the task. Then, participants were allowed to perform the next task on the test device.

When the participant accomplished all tasks on one ventilator, they would move to the next ventilator to repeat the tasks. Each participant was asked to perform test tasks continuously on the 3 test ventilators. They were permitted rest whenever needed.

Data collection

When a participant completed a task, the task completion time and pupil diameter data would be recorded, and an expert would identify completion status of the task: success or failure. After completion of all tests at one tested ventilator, the subjective evaluation questionnaire would complete

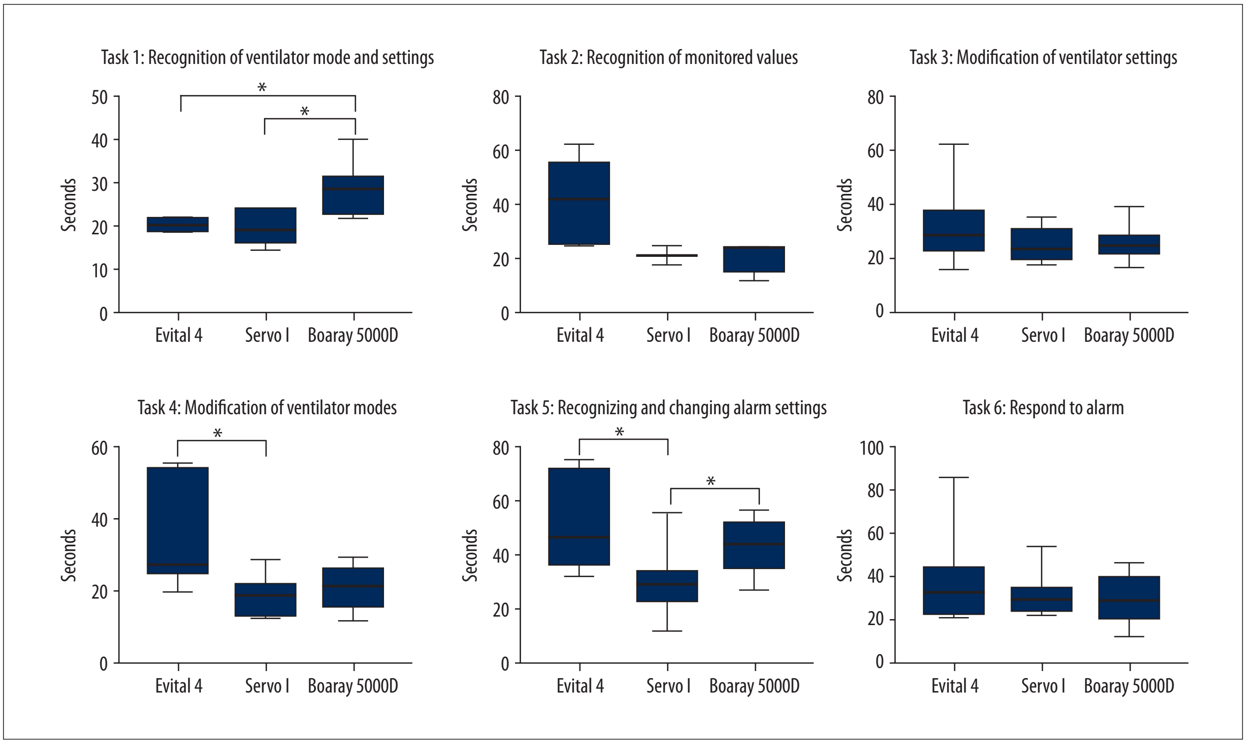

Figure 1. Box-plot showing the completion time needed for each ventilator for performance of the 6 tasks. Task 1: recognition of ventilator mode and settings, Task 4: modification of ventilator modes, and Task 5: recognizing and changing alarm settings showed statistical statistically significant differences among the 3 ventilators (* P<0.05).

by participants to evaluate the ergonomics of the ventilator's user-interfaces. Finally, an interview with participants would be conducted to collected qualitative data regarding ergonomic design of the ventilator user-interface.

Statistics

Data values are expressed as the mean ±SD. Task completion time, pupil data, and subjective evaluation data were performed using Friedman non-parametric test to compare differences among ventilators. Post-hoc multiple comparison tests for any 2 ventilators were performed using the Dunn-Bonferroni test [35]. P<0.05 was considered statistically significant. Analyses were performed using statistics software SPSS 20 (IBM Corporation, Armonk, New York).

Results

Task completion time and failures

Figure 1 shows the total time needed to complete the 6 tasks for each ventilator. Recognition of ventilator mode and settings, modification of ventilator modes, and recognizing and changing alarm settings were different among the 3 ventilators (Supplementary Table 2). Post-hoc comparison analysis was made using the Dunn-Bonferroni test (Supplementary Table 3). This showed that participants needed more time to complete the task of recognition of ventilator mode and settings on the Boaray 5000D than on the Evital 4 (P=0.028) and Servo I (P=0.028). For task of modification of ventilator modes, participants needed less to complete the task on the Servo I than Evital 4 (P=0.007). Participants needed less time to complete the task of recognizing and changing alarm settings on the Servo I than on the Evital 4 (P=0.011) or on the Boaray 5000D (P=0.042).

Table 1 presents the task failures for each ventilator for 6 tasks; a mean of 15(2) tasks failures was observed for each machine (15.6%). There were no significant differences in task failures among 3 machines (P=0.415). Task failure most often occurred for recognition of monitored values (45.7% of all failures), followed by recognition of ventilator mode and settings (26.1%), and respond to alarm (10.9%).

Pupil diameter change from baseline

Table 2 displays the change from baseline in the participant pupil diameter when accomplishing 6 tasks on 3 ventilators. For only task 2, the pupil diameter change did not show significant differences among ventilators for in the task of recognition of monitored values (P=0.233). For the other tasks,

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 1. Task failures.

Tasks	Ventilators			
	Evital 4	Servio I	Boaray 5000D	Total
Task 1: Recognition of ventilator mode and settings	3	5	4	12
Task 2: Recognition of monitored values	6	9	6	21
Task 3: Modification of ventilator settings	1	1	1	3
Task 4: Modification of ventilator modes	0	0	2	2
Task 5: Recognizing and changing alarm settings	1	1	1	3
Task 6: Respond to alarm	4	1	0	5
Total	**15**	**17**	**14**	**46**

Table 2. Change in pupil diameter from baseline during performing six test tasks in ventilators.

Tasks	Evital 4	Servo I	Boaray 5000D	p
	Mean ±SD	Mean ±SD	Mean ±SD	
Task 1: Recognition of ventilator mode and settings (mm)	0.153±0.032	0.062±0.025	0.179±0.035	0.009*
Task 2: Recognition of monitored values (mm)	0.088±0.011	0.055±0.007	0.066±0.015	0.223
Task 3: Modification of ventilator settings (mm)	0.165±0.013	0.133±0.013	0.130±0.012	0.004*
Task 4: Modification of ventilator modes (mm)	0.178±0.038	0.121±0.041	0.126±0.053	0.023*
Task 5: Recognizing and changing alarm settings (mm)	0.189±0.021	0.136±0.057	0.143±0.031	0.003*
Task 6: Respond to alarm (mm)	0.196±0.028	0.160±0.021	0.135±0.022	0.004*

* Statistically significant results.

significantly different changes in participants' pupil size from baseline were found among 3 machines ($P<0.05$).

Post-hoc multiple comparisons of change in pupil diameter from baseline between ventilators were analyzed using the Dunn-Bonferroni test (Table 3). After Bonferroni correction, 8 out of 15 comparisons were statistically significant.

The largest change in pupil diameter from baseline was recorded while accomplishing tasks on the Evital 4, compared with the other ventilators (except for task 1). Furthermore, the participants' pupil diameter increased during performance tasks in the 3 ventilators.

The average slope of pupil diameter changed from baseline over time

Table 4 shows the average slope of pupil diameter change from baseline over time during performance of the 6 tasks; we found that the average slopes of pupil diameter change from baseline over time were significant for task 1, 3, 4, 5, 6 among the 3 machines ($P<0.05$).

Post-hoc multiple comparisons of the average slope of pupil diameter change from baseline over time between ventilators is displayed in Table 5. After Bonferroni correction, 8 out of 15 comparisons were statistically significant.

Tables 4 and 5 display the average slope of pupil diameter change from baseline over time shown. Performance of the tasks (except for task 1 and task 2) on the Evital 4 were more difficult than on the Servo I or Boaray 5000D, as shown by pupil dilation. The Servo I was easy for participants to perform the tasks, shown by pupil relaxation. This was similar for the Boaray 5000D (except for task 1).

Correlation analysis between task completion time and the change in pupil diameter

The results of correlation analysis between task completion time and the change in pupil diameter from baseline are displayed

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 3. Post hoc multiple comparison of change in pupil diameter from baseline between ventilators.

	Smaller change in pupil diameter	Test statistic	S.E.	Bonferroni correction for p-value
Task 1: Recognition of ventilator mode and settings				
Servo I – Evital 4		1.333	0.577	0.063
Servo I – Boaray 5000D	Servo I	1.667	2.887	0.012*
Evital 4 – Boaray 5000D		0.333	0.577	1.000
Task 3: Modification of ventilator settings				
Servo I – Evital 4	Servo I	1.250	0.447	0.016*
Servo I – Boaray 5000D		0.100	0.447	1.000
Boaray 5000D – Evital 4	Boaray 5000D	−1.150	0.447	0.030*
Task 4: Modification of ventilator modes				
Servo I – Evital 4	Servo I	1.167	0.471	0.040*
Servo I – Boaray 5000D		0.167	0.471	1.000
Boaray 5000D – Evital 4		−1.000	0.471	0.102
Task 5: Recognizing and changing alarm settings				
Servo I – Evital 4	Servo I	1.450	0.447	0.004*
Servo I – Boaray 5000D		0.350	0.447	1.000
Boaray 5000D – Evital 4	Boaray 5000D	−1.100	0.447	0.042*
Task 6: Respond to alarm				
Servo I – Evital 4	Servo I	1.286	0.535	0.048*
Boaray 5000D – Servo I		−0.429	0.535	1.000
Boaray 5000D – Evital 4	Boaray 5000D	−1.714	0.535	0.004*

* Statistically significant results.

Table 4. The average slope of pupil diameter change from baseline over time.

Tasks	Evital 4	Servo I	Boaray 5000D	p
	Mean ±SD	Mean ±SD	Mean ±SD	
Task 1: Recognition of ventilator mode and settings (mm/sec)	−0.021±0.021	−0.019±0.021	0.010±0.012	0.011*
Task 2: Recognition of monitored values (mm/sec)	−0.002±0.004	−0.009±0.005	−0.012±0.012	1.000
Task 3: Modification of ventilator settings (mm/sec)	0.020±0.027	−0.016±0.012	−0.013±0.015	0.025*
Task 4: Modification of ventilator modes (mm/sec)	0.015±0.023	−0.023±0.020	−0.010±0.007	0.008*
Task 5: Recognizing and changing alarm settings (mm/sec)	0.010±0.014	−0.030±0.022	−0.010±0.007	<0.001*
Task 6: Respond to alarm (mm/sec)	0.034±0.036	−0.013±0.004	−0.015±0.023	0.004*

* Statistically significant results.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 5. Post hoc multiple comparisons of the slope of the pupil diameter change from baseline over time between ventilators.

	Smaller the slope	Test statistic	S.E.	Bonferroni correction for p-value
Task 1: Recognition of ventilator mode and settings				
Servo I – Boaray 5000D	Servo I	1.500	0.577	0.028*
Evital 4 – Boaray 5000D	Evital 4	1.500	0.577	0.028*
Evital 4 – Servo I		0.000	0.577	1.000
Task 3: Modification of ventilator settings				
Servo I – Evital 4	Servo I	1.100	0.447	0.042*
Servo I – Boaray 5000D		0.100	0.447	1.000
Boaray 5000D – Evital 4		–1.000	0.447	0.076
Task 4: Modification of ventilator modes				
Servo I – Evital 4	Servo I	1.444	0.471	0.007*
Servo I – Boaray 5000D		0.556	0.471	0.716
Boaray 5000D – Evital 4		–0.889	0.471	0.178
Task 5: Recognizing and changing alarm settings				
Servo I – Evital 4	Servo I	1.800	0.447	< 0.001*
Servo I –Boaray 5000D		0.600	0.447	0.539
Boaray 5000D – Evital 4	Boaray 5000D	–1.200	0.447	0.022*
Task 6: Respond to alarm				
Servo I – Evital 4	Servo I	1.714	0.535	0.004*
Servo I – Boaray 5000D		0.429	0.535	1.000
Boaray 5000D – Evital 4	Boaray 5000D	–1.286	0.535	0.048*

* Statistically significant results.

in Table 6. We found that the task completion time and the change in pupil diameter from baseline during different tasks among ventilators had several correlations. For the Evital 4, during task 1, the task completion time was positively correlated with the change in pupil diameter (r=0.830, *P*=0.011); for task 2, the task completion time was negatively correlated with the change in pupil diameter (r=–0.894, *P*=0.041). For the Servo I, in task 1, 2, 4, 5, 6, the task completion time were significantly positively correlated with the change in pupil diameter (r=0.985, *P*=0.000; r=1.000, *P*=0.000; r=0.482, *P*=0.000; r=0.640, *P*=0.0046; and r=0.874, *P*=0.001, respectively). For the Boaray 5000D, the task completion time was significantly positively correlated with the change in pupil diameter in task 4, 5 (r=0.675, *P*=0.032 and r=0.916, *P*=0.001, respectively).

Table 6 shows that the task completion time was better positively correlated with the change in pupil diameter for participants performing tasks on the Servo I (except task 3). The Evital 4 and Boaray 5000D only showed correlations in several tasks.

Subjective evaluation

Table 7 shows the results of subjective evaluation. Statistically significant differences were found in friendliness of user-interface (*P* = 0.005), information display friendliness of user-interface (*P*<0.001), and safety of user-interface (*P* = 0.001). Post-hoc comparisons with Bonferroni correction were performed with these data (Supplementary Table 4). Participants considered the Servo I to be more user-friendliness than the Evital 4 (*P*=0.018); for information display, the Servo I was better than the Evital 4 (*P*=0.001); for safety of user-interface, the Boaray 5000D was safer than the Evital 4 (*P*=0.013). The Evital 4 got the worst results in the post study subjective evaluation.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 6. Correlation analysis between the task completion time and the change in pupil diameter from baseline during performing the six tasks.

Task	Ventilator					
	Evital 4		Servo I		Boaray 5000D	
	Coefficient	p-Value	Coefficient	p-Value	Coefficient	p-Value
Task 1: Recognition of ventilator mode and settings	0.830	0.011*	0.985	<0.001*	–0.256	0.579
Task 2: Recognition of monitored values	–0.894	0.041*	1.000	<0.001*	0.667	0.219
Task 3: Modification of ventilator settings	0.164	0.651	0.482	0.159	0.675	0.032*
Task 4: Modification of ventilator modes	–0.311	0.353	0.975	<0.001*	0.916	0.001*
Task 5: Recognizing and changing alarm settings	–0.116	0.749	0.640	0.046*	–0.018	0.960
Task 6: Respond to alarm	–0.649	0.115	0.874	0.001*	–0.233	0.490

* Statistically significant results.

Table 7. Subjective evaluation.

	Evital 4	Servo I	Boaray 5000D	p
	Mean ±SD	Mean ±SD	Mean ±SD	
Friendliness of User-Interface: 1 (very bad) to 5 (very good), Mean ±SD	3.250±0.754	4.417±0.669	3.750±0.754	0.005*
Information Display Friendliness of User-Interface: 1 (very bad) to 5 (very good), Mean ±SD	3.750±0.866	4.917±0.289	4.500±0.674	<0.001*
Safety of User-Interface: 1 (certainly not) to 5 (certainly), Mean ±SD	3.417±0.900	4.417±0.669	4.667±0.492	0.001*

* Statistically significant results.

Discussion

Our usability study of ventilator user-interface showed that most ventilators now used in China have poor ergonomics, with high task error rates (21.2% to 35.0%) and longer task completion times than those reported in published studies [17–21]; this can lead to serious consequences in emergency situations. These can be illustrated by several examples observed in our study. First, the terms of ventilation parameters are not uniform among ventilators. For example, on the ventilation parameters settings page, the fraction of inspired oxygen was listed as"O_2[%] "on the Evita 4,"O_2" on the Servo I and "$FiO_{2(\%)}$" on the Boaray 5000D. This can cause confusion in emergency situations. Second, the adjustment method of ventilation mode and parameters are different among ventilators. For the Evital 4, the adjustment of mode and parameters only requires to operation of a knob in the bottom right corner of the screen, with confirmation of the change by pressing the knob. However, in the Boaray 5000D and Servo I, in addition to the operation common to the Evital 4, one needs to click an "accept" button on the screen, a task most participants found easy to forget. The difference in design of the user-interface resulted in unnecessary operational failures and increased the task completion times; our study showed worse results compared with other studies [17–20,22]. These problems have also been reported by several published studies. Marjanovic and L'Her [20] found that non-uniform terminology acronyms resulted in mode recognition; Uzawa et al. [17] showed that differences in terminology caused unnecessary confusion for users and increased failures and task completion times; Templier et al. [18] also reached similar conclusions.

There are few published studies applying eye-tracking data to evaluate the user-interface of ventilators in usability testing. Our usability study demonstrated that pupil diameter can be used as a standardized assessment indicator for ergonomic evaluation of ventilator user-interfaces [26]. In this study, we found statistically significant differences on 2 pupil diameter variables (pupil diameter change and the slope of pupil diameter change over time). According to published studies, pupil diameter was a sensitive index of the task difficultly and cognitive demands, with increased task difficultly and cognitive demands indicated by an increase in pupil diameter change from baseline [26,29–31,36]. The mean changes in pupil diameter and the slope of pupil diameter changes over time during six tasks among three ventilators are shown in Tables 2 and 4, and statistically significant differences were found in task 1, 3, 4, 5, 6. For the Servo I, the data from Table 3 show

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

that the change of pupil diameter was smaller than the Boaray 5000D in task 1 (*P*=0.012), and smaller than the Evital 4 in task 3, 4, 5, 6 (*P*=0.016; *P*=0.040; *P*=0.004; *P*=0.048, respectively). These results indicate that the Servo I outperformed the Boaray 5000D in task 1, and outperformed the Evital 4 in task 3, 4, 5, 6. These conclusions were further supported by the result of the slope of pupil diameter changes over time. The data from Tables 4 and 5 show that the pupil diameter increased significantly for Boaray 5000D while started to decrease for Servo I during task 1 (*P*=0.028). The pupil diameter enlarged for Boaray 5000D and shrank for Servo I means that the mental workload demanded by Servo I to complete task 1 is lower than Boaray 5000D. For task 3, 4, 5, 6, we can also reach similar conclusions that the mental workload demanded by Servo I to complete task is lower than Evital 4. For Boaray 5000D, the data from Tables 3 and 5 also show that it outperformed Evital 4 in task 5 and task 6. The Evital 4 outperformed Boaray 5000D only in task 1 (Table 3). As for the 2 pupil diameter variables, Evital 4 got a bad performance for the tested tasks. These conclusions indicate that the pupil diameter change and the average slope of pupil diameter change from baseline over time can be used as standardized evaluation methods together with task completion time and task failures to evaluate the ergonomics of the user-interface for ventilators.

Furthermore, the correlation between task completion time and the change in pupil diameter was analyzed and is shown in Table 6. Our study results showed that several significant correlations were found among the 3 ventilators during completing the 6 tasks. The results suggest that there might be a correlation between task completion time and the change in pupil diameter. A more detailed study should be explored in future to analyze this correlation. Eye-tracking data were supported by participants' subjective evaluation, in which Evital 4 showed the worst scores in terms of the friendliness of the user-interface (3.250±0.754, *P*=0.005), information display friendliness of user-interface (3.750±0.866, *P*<0.001), and safety of user-interface (3.417±0.900, *P*=0.001). In similar studies [17–20], ergonomic evaluation heavily depended on task completion time, task failure, subjective evaluation, and mental work load, which are ignored participants' physiological parameter to evaluate of the ergonomics of the user-interface. The pupil diameter measurements to assess ergonomic of user interface are new method in the ICU field. Compared to subjective measurements, these eye-tracking data that permit the estimation of the workload caused by a ventilator's user interface and an indirect evaluation of the user interface's ergonomic. Therefore, pupil diameter can be a promising tool for the assessment of medical device user-interfaces in usability testing.

Several new generation ventilators have been developed by manufacturers. However, according to a recently published study, the usability problems we found in tested ventilators remain persist in the new generation ventilators, including absence of a consistent terminology among ventilators, user interface simplification, rationalization, and important settings display [26]. Our study results are important for manufacturers to improve the ergonomics of ventilator user interfaces.

Limitations

Some limitations of this study must be recognized. First, the participants of this study only represent respiratory therapists, thus the results cannot be directly generalized to other user groups. Second, only evaluating three ventilators may be insufficient. However, the 3 ventilators are widely used in our region medical institutions, and all of them were available for testing. Third, the tested ventilators were not the new generation ventilators, but the tested ventilators were most used in our local medical institutions and were available for study. Our study intent was to provide a method to evaluate ventilator ergonomics. Finally, the test ventilators have more functions than were tested in our study.

Conclusions

The usability study of 16 respiratory therapists proved that most ventilators used in China have ergonomic problems, which also be reported by published studies [37–40]. Therefore, it may be better to optimize the design of user-interfaces of ventilators to avoid these ergonomic problems leading to use errors. Furthermore, this study used eye tracking data pupil diameter as a new tool to evaluate the ergonomics of ventilator user-interfaces, returning positive results. Further study may deepen analysis of eye tracking data to reflect the user's real characteristics, making the user-interface of ventilator more adaptable to Chinese end-users.

Acknowledgements

The authors would like to acknowledge the help of all respiratory therapists from Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, who made this study possible.

Conflicts of interests

None.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Supplementary Materials

Tasks to accomplish

The tasks to accomplish were the following:

1. Recognition of ventilator mode and settings

With the tested ventilator turned on and running in a preset ventilation mode, the participant had to answer the following questions: a) whether the ventilator was in VC-IMV or PC-CSV; b) what were the settings in current ventilation mode [VC-IMV: inspired oxygen fraction (FIO_2), tidal volume (VT), respiratory rate (RR), positive end-expiratory pressure (PEEP); PC-CSV: inspired oxygen fraction (FIO_2), respiratory rate (RR), positive end-expiratory pressure (PEEP) and inspiratory pressure (P_{insp})]. The stop signal was given when participant had answered the all questions.

The first ventilator was in mode of VC-IMV, then PC-CSV and VC-IMV ventilation mode were alternated on each subsequent ventilator. The ventilation settings were: VC-IMV: FIO_2 0.6, VT 500 mL, RR 15/min, PEEP 5 cm H_2O; PC-CSV: FIO_2 0.3, RR 18/min, PEEP 3 cm H_2O, P_{insp} 15 cm H_2O.

2. Recognition of monitored values

With the tested ventilator turned on and running in a preset ventilation mode, participant had to inform the testers the monitored values: VC-IMV [plateau pressure (P_{plat}), peak inspiratory pressure (P_{peak}), minute volume (MV), expired tidal volume (V_{Te})]; PC-CSV [minute volume (MV), respiratory rate (RR), positive end-expiratory pressure (PEEP) and tidal volume (VT)]. The stop signal was given when participant had reported the all required monitored values.

3. Modification of ventilator settings

With the tested ventilator turned on and running in a preset ventilation mode, the participant had to change the ventilation settings. The ventilation settings changed values were: VC-IMV: FIO_2 0.8, VT 600 mL, RR 20/min, PEEP 10 cm H_2O; PC-CSV: FIO2 0.5, RR 12/min, PEEP 8 cm H_2O, P_{insp} 10 cm H_2O. The stop signal was given when all required setting values were adjusted and activated.

4. Modification of ventilator modes

With the tested ventilator turned on and running in a preset ventilation mode, participant had to change from VC-IMV to PC-CSV or PC-CSV to VC-IMV. The first ventilator was in mode of VC-IMV, then PC-CSV and VC-IMV ventilation mode were alternated on each subsequent ventilator. The stop signal was given with the first insufflation of new mode.

5. Recognizing and changing alarm settings

With the tested ventilator turned on and running in a preset ventilation mode, participant had to inform tester several alarm settings: minute volume (MV), respiratory rate (RR), airway pressure (P_{aw}). After participant reported the values of alarm settings, participant had to change the value of alarm settings form the present level to the requiring level. The stop signal was given as soon as the adjustment was activated.

6. Respond to alarm

With the tested ventilator turned on and running in a preset ventilation mode, the tester changed one alarm setting to trigger an alarm, the participant had to stop alarm, report the alarm content, adjust alarm to predefined values, and reset the alarm. The stop signal was given when alarm values had been adjusted to required levels. In this study, the alarms were: low pressure, high tidal volume and apnea, alternated in that order between ventilators.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Supplementary Table 1. Randomisation table for device's testings.

Participants number	Ventilator type		
1	Evital 4	Servo I	Boaray 5000D
2	Evital 4	Boaray 5000D	Servo I
3	Servo I	Boaray 5000D	Evital 4
4	Servo I	Evital 4	Boaray 5000D
5	Boaray 5000D	Servo I	Evital 4
6	Boaray 5000D	Evital 4	Servo I
7	Evital 4	Servo I	Boaray 5000D
8	Servo I	Boaray 5000D	Evital 4
9	Boaray 5000D	Servo I	Evital 4
10	Evital 4	Boaray 5000D	Servo I
11	Servo I	Evital 4	Boaray 5000D
12	Evital 4	Servo I	Boaray 5000D
13	Servo I	Boaray 5000D	Evital 4
14	Boaray 5000D	Servo I	Evital 4
15	Evital 4	Boaray 5000D	Servo I
16	Servo I	Evital 4	Boaray 5000D

This table details the randomization table for devices' testing by the respiratory therapists.

Supplementary Table 2. Task completion time.

Tasks	Evital 4	Servo I	Boaray 5000D	p
	Mean ±SD	Mean ±SD	Mean ±SD	
Task 1: Recognition of ventilator mode and settings	20.400±1.317	19.650±3.725	28.386±6.116	0.011*
Task 2: Recognition of monitored values	40.860±15.468	21.500±4.950	20.740±4.979	0.223
Task 3: Modification of ventilator settings	31.540±13.057	25.250±6.128	25.940±6.293	0.368
Task 4: Modification of ventilator modes	21.400±5.897	18.327±5.252	21.400±0.053	0.008*
Task 5: Recognizing and changing alarm settings	52.860±17.599	29.800±11.827	43.030±9.928	0.007*
Task 6: Respond to alarm	38.414±21.702	30.700±9.090	28.109±10.664	0.368

* Statistically significant results.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Supplementary Table 3. Multiple comparison of task completion time.

	Lower the time	Test statistic	S.E.	Bonferroni correction for p-value
Task 1: Recognition of ventilator mode and settings				
Servo I – Boaray 5000D	Servo I	1.500	0.577	0.028*
Evital 4 – Boaray 5000D	Evital 4	1.500	0.577	0.028*
Evital 4 – Servo I		0.000	0.577	1.000
Task 4: Modification of ventilator modes				
Servo I – Evital 4	Servo I	1.444	0.471	0.007*
Servo I – Boaray 5000D		0.889	0.471	0.178
Boaray 5000D – Evital 4		–0.556	0.471	0.716
Task 5: Recognizing and changing alarm settings				
Servo I – Evital 4	Servo I	1.300	0.447	0.011*
Servo I – Boaray 5000D	Servo I	1.100	0.447	0.042*
Boaray 5000D – Evital 4		–0.200	0.447	1.000

* Statistically significant results.

Supplementary Table 4. Multiple comparison of subjective evaluation.

	Lower the slope	Test statistic	S.E.	Bonferroni correction for p-value
Friendliness of user-interface				
Boaray 5000D – Servo I		–0.625	0.408	0.377
Evital 4 – Boaray 5000D		0.500	0.408	0.662
Evital 4 – Servo I	Servo I	–1.125	0.408	0.018*
Information display friendliness of user-interface				
Evital 4 – Servo I	Servo I	–1.458	0.408	0.001*
Boaray 5000D – Servo I		–0.542	0.408	0.554
Evital 4 – Boaray 5000D		0.917	0.408	0.074
Safety of user-interface				
Evital 4 – Servo I		–0.833	0.408	0.124
Servo I – Boaray 5000D		0.333	0.408	1.000
Evital 4 – Boaray 5000D	Boaray 5000D	1.167	0.408	0.013*

* Statistically significant results.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

References:

1. Auriant I, Reignier J, Pibarot ML et al: Critical incidents related to invasive mechanical ventilation in the ICU: Preliminary descriptive study. Intens Care Med, 2002; 28(4): 452–58
2. Chatwin M, Heather S, Hanak A et al: Analysis of home support and ventilator malfunction in 1,211 ventilator-dependent patients. Eur Respir J, 2010; 35(2): 310–16
3. ECRI. Top 10 health technology hazards for 2016. Emergency Care Research Institute, 2016
4. Bracco D, Favre J-B, Bissonnette B et al: Human errors in a multidisciplinary intensive care unit: A 1-year prospective study. Intens Care Med, 2001; 27(1): 137–45
5. Gopher D, Olin M, Badihi Y et al: The nature and causes of human errors in a medical intensive care unit. Human Factors & Ergonomics Society Annual Meeting Proceedings, 1989; 33(15): 956–60
6. Wright D, Mackenzie SJ, Buchan I et al: Critical incidents in the intensive therapy unit. Lancet, 1991; 338(8768): 676–78
7. Gajic O, Frutos-Vivar F, Esteban A et al: Ventilator settings as a risk factor for acute respiratory distress syndrome in mechanically ventilated patients. Intens Care Med, 2005; 31(7): 922–26
8. Donchin Y, Seagull FJ: The hostile environment of the intensive care unit. Curr Opin Crit Care, 2002; 8(4): 316–20
9. Crowley, John J, Kaye, Ronald D: Identifying and understanding medical device use errors. J Clin Eng, 2002; 27(3): 188–93
10. Bogner MS: There is More to Error in Healthcare than the Care Provider. Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting; 2005 May 29; Florida, USA
11. Israelski EW, Muto WH: Human factors risk management as a way to improve medical device safety: A case study of the therac 25 radiation therapy system. Jt Comm J Qual Saf, 2004; 30(12): 689–95
12. IEC. IEC 62366-1 Medical Devices – PART 1: Application of Usability Eegineering to Medical Devices. Geneva: International Electrotechnical Commission; 2015
13. IEC. Medical Electrical Equipment – Part 1–6: General Requirements For Basic Safety And Essential Performance. Geneva: Usability International Electrotechnical Commission; 2013
14. FDA (US). Applying Human Factors and Usability Engineering to Medical Devices. U.S. Food and Drug Administration. Washington: Food and Drug Administration; 2016
15. Wiklund ME, Kendler J, Strochlic AY: Usability testing of medical devices. Florida: Taylor & Francis; 2011
16. Nielsen J: Usability engineering. Boston: Academic Press; 1993
17. Uzawa Y, Yamada Y, Suzukawa M: Evaluation of the user interface simplicity in the modern generation of mechanical ventilators. Resp Care, 2008; 53(3): 329–37
18. Templier F, Miroux P, Dolveck F et al: Evaluation of the ventilator-user interface of 2 new advanced compact transport ventilators. Resp Care, 2007; 52(12): 1701–9
19. Gonzalezbermejo J, Laplanche V, Husseini FE et al: Evaluation of the user-friendliness of 11 home mechanical ventilators. Eur Respir J, 2006; 27(6): 1236–43
20. Marjanovic N, L'Her E: A comprehensive approach for the ergonomic evaluation of 13 emergency and transport ventilators. Resp Care, 2016; 61(5): 632–39
21. Morita PP, Weinstein PB, Flewwelling CJ et al: The usability of ventilators: A comparative evaluation of use safety and user experience. Crit Care, 2016; 20(1): 263–72
22. Vignaux L, Tassaux D, Jolliet P: Evaluation of the user-friendliness of seven new generation intensive care ventilators. Intens Care Med, 2009; 35(10): 1687–91
23. Spaeth J, Schweizer T, Schmutz A et al: Comparative usability of modern anaesthesia ventilators: A human factors study. Brit J Anaesth, 2017; 119(5): 1000–8
24. Zhang JY, Liu SL, Feng QM et al: Correlative evaluation of mental and physical workload of laparoscopic surgeons based on surface electromyography and eye-tracking signals. Sci Rep-UK, 2017; 7 (1): 11095
25. Browning M, Cooper S, Cant R et al: The use and limits of eye-tracking in high-fidelity clinical scenarios: A pilot study. Int Emerg Nurs, 2015; 25: 43–47
26. Marjanovic NS, De Simone A, Jegou G, L'Her E: A new global and comprehensive model for ICU ventilator performances evaluation. Ann Intensive Care, 2017; 7(1): 68–81
27. Richstone L, Schwartz MJ, Seideman C et al: Eye metrics as an objective assessment of surgical skill. Ann Surg, 2010; 252(1): 177–82
28. Bradshaw JL: Pupil size and problem solving. Q J Exp Psychol, 1986; 20(2): 116–22
29. Gable TM, Kun AL, Walker BN, Winton RJ: Comparing heart rate and pupil size as objective measures of workload in the driving context: initial look. Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications; 2015 September 01–03; Nottingham, United Kingdom
30. Reiner M, Gelfeld TM: Estimating mental workload through event-related fluctuations of pupil area during a task in a virtual world. Int J Psychophysiol, 2014; 93(1): 38–44
31. Liu Z, Xing B, Qianxiang Z, Zhang X: The Experiment Research of Pupil Change for the Evaluation of Mental Workload. Proceedings of the18th Internal Conference on Human-Computer Interaction;2016 July 17–22; Tronoto, Canada
32. Hwang W, Salvendy G: Number of people required for usability evaluation: The 10±2 rule. Communications of the ACM, 2010; 53(5): 130–33
33. Virzi RA: Refining the test phase of usability evaluation: How many subjects is enough? Hum Factors, 1992; 34: 457–68
34. Jiang X, Zheng B, Tien G, Atkins MS: Pupil response to precision in surgical task execution. Stud Health Technol Inform, 2013; 184: 210–14
35. Dunn OJ: Multiple comparison using RANK sums. Technometrics, 1964; 6(3): 241–52
36. Verney SP, Granholm E, Dionisio DP: Pupillary responses and processing resources on the visual backward masking task. Psychophysiology, 2010; 38(1): 76–83
37. Chen Q, Sha Y: [Safety risk assessment of ventilator.] Chin Med Equip J, 2015; 36(06): 118–20 [in Chinese]
38. Jiang Y, Peng W, Ma L: [Actuality analysis and safety risk management measures of ventilator use.] Chin Med Devices, 2012; 27(02): 71–73 [in Chinese]
39. Zhang L, Zhao Y, Dong F: Medical device adverse events monitoring and risk analysis on ventilators. Chin J Pharma, 2016; 13(08): 503–4 [in Chinese]
40. Yan Q, Wu J: [The clinical engineering strategy for reducing application risk of ventilator.] Chin Med Equip, 2018; 15(03): 36–39 [in Chinese]

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]