Retention of military combat lifesaving skills during six months following classroom-style and individualized-style initial training

Landman, Annemarie; de Vries, Daný; Binsch, Olaf

DOI
10.1080/08995605.2022.2144034

Publication date
2022

Document Version
Final published version

Published in
Military Psychology

Military Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hmlp20

Retention of military combat lifesaving skills during six months following classroom-style and individualized-style initial training

Annemarie Landman, Daný de Vries & Olaf Binsch

MILITARY PSYCHOLOGY
https://doi.org/10.1080/08995605.2022.2144034

Retention of military combat lifesaving skills during six months following classroom-style and individualized-style initial training

Annemarie Landman [a,b], Daný de Vries[a], and Olaf Binsch [a,c]

[a]Netherlands Organization for Applied Scientific Research (TNO), The Netherlands; [b]Delft University of Technology, Delft, The Netherlands; [c]Health Department of the Royal Netherlands Army, Netherlands Ministry of Defense, Utrecht, The Netherlands

ABSTRACT

The current study was performed to obtain insight into the retention of combat lifesaving (CLS) skills after initial training and to compare a more individualized-style training with a more classroom-style training. We measured performance at 0 month, 2 months, and 6 months after initial training in 40 CLSers (17 individualized, 23 classroom). Each test consisted of two 20-minute scenarios with a medical mannequin to simulate combat injuries. An instructor scored the actions, which were divided into critical and non-critical by medical experts. We also measured the speed of performing the protocol and perceived mental effort and anxiety. There were no differences between the groups in critical actions. The full sample made on average almost six critical errors per scenario at 6 months. However, on non-critical actions, the individualized group scored better at 0 month. The individualized group also performed the protocol faster at each test. The classroom group reported an increase in mental effort and anxiety at subsequent tests, while the individualized group did not. Based on the high number of critical errors at 6 months, and on the drop-off in performance at 2 months, we advise that extra refresher training is organized within 2 months after initial training to improve retention further down the line.

ARTICLE HISTORY
Received 15 April 2022
Accepted 25 October 2022

KEYWORDS
Currency; education; medic; personalized; skill decay

What is the public significance of this article?— We longitudinally investigated the skill decay in military combat lifesavers for 6 months following initial training and compared a more classroom style of initial training with a more individualized style. Some advantages were found of individualized training. The high amount if critical errors at 6 months for both styles of training led us to conclude that early refresher training for combat lifesaving skills or similar skills is advised.

Introduction

Professionals in safety-critical occupations, such as the military, the fire service, or the police, are entrusted to deal with the most extraordinary circumstances. The skills required for these circumstances are used infrequently in operational environments and must therefore be periodically trained and assessed in safe, simulated settings (Sullivan et al., 2019). Such refresher training is usually followed by a test to ensure that the level of competence is acceptable (i.e., "current") and that the skill can be performed well under pressure (see Oprins et al., 2019). Unfortunately, such tests do not provide insight into the skill decay, because they are preceded by the refresher training. Without such insight, the frequency and timing of refresher training remains based on subjective experience of legislators instead of on operational requirements and evidence.

Skill decay refers to a loss of acquired skills or knowledge after a period of nonuse (Arthur et al., 1998, p. 58; Linde et al., 2018). Nonuse may seriously affect performance in particular in high-risk professions, as acute stress impairs one's ability to recall declarative knowledge while well-learned procedures remain relatively robust (Hancock & Szalma, 2008). The opposite of skill decay is called retention, which is the extent to which previously learned skills are maintained despite nonuse. Longitudinal empirical data on domain-specific skills are necessary to determine the required content and frequency of refresher training to ensure an acceptable level of retention. In the current study, we aim to provide such data on retention of medical skills in the military domain, that is, the skills of combat lifesaving (CLS). CLSers are the first responders to medical emergencies for military personnel as well as civilians.

During initial training, CLSers learn to treat afflictions that are typical for military combat, such as shot

CONTACT Olaf Binsch olaf.binsch@tno.nl Human Performance, Netherlands Organization for Applied Scientific Research (TNO); Kampweg 55, 3769 DE Soesterberg, The Netherlands.

wounds, blast injuries, or blunt force injuries. For treatment, a mnemonic-based protocol is used to help remember the priority-ordered diagnostic and treatment actions under stress (see also, Landman et al., 2020; Wisher et al., 1999). At the time of the current study, the Dutch military used the <C> *ABCDE* protocol, which stands for: Catastrophic hemorrhage, Airway, Breathing, Circulation, Disability, and Environment. Since then, they have switched to the *MARCH* protocol to be more in line with Tactical Combat Casualty Care (TCCC) procedures used in other NATO countries. The protocols are highly similar, as *MARCH* stands for Massive bleeding, Airway, Respiration, Circulation, Hypothermia (Center for Army Lessons Learned, 2017). CLSers also learn the underlying reasons for many of the actions, so that they can make decisions about the allocation of time, effort, and attention. CLS tasks thus consist of motor skills (e.g., applying a tourniquet), procedural skills (e.g., remembering the several actions falling under the letter *A*) and judgment skills (e.g., combining different cues to determine the seriousness of an issue). If combat deployment is planned in advance, this is usually reason for extra, more intensive, refresher courses to ensure currency. However, deployment cannot always be anticipated, and sudden emergencies such as natural disasters may also call for CLS deployment. Furthermore, CLSers are always depended upon to provide first-aid when accidents occur during military training. It is therefore important to continuously minimize skill decay in this group.

There are currently no empirical longitudinal data on the retention of CLS skills, and applying recent findings from literature of medical, military, or other related domains is problematic due to the high variety in findings. A systematic review of the retention of skills in safety-critical professions mainly in the medical domain showed that the skill decay at 6 months since initial training varied between ca. 30% and 60% decay and at 12 months between 50% and 100% (Figure 2(b) in Vlasblom et al., 2020). The authors stated that they excluded outlier results from this analysis, which were possibly caused by exceptional training quality or low performance at initial training. Skill decay also varied in a review of multi-day training courses on medical, military, and offshore safety by Sanli and Carnahan (2018). They concluded that retention can be expected for 6 months at best, with skills (i.e., practiced abilities) decaying more quickly than declarative knowledge, complex tasks being forgotten more quickly than simple tasks, and higher initial skill level of trainees being related to higher retention. The variance in skill decay was similarly high in advanced life support in a systematic review by Yang et al. (2012), with some studies reporting minor skill decay of 3–7% at 6 months and other studies reporting that 40–86% of healthcare providers performed inadequately at 1 year. They conclude that a decline starts to occur between 6 months and 1 year, although this estimate was also based on studies where subjects had clinical experience prior or after initial training. With regard to cardiopulmonary resuscitation (CPR) skills in nurses, Hamilton (2005) found that skills and declarative knowledge decayed between 3 and 6 months after training, with motor skills decaying more quickly than knowledge. Detectable declines in motor skills started already at 2 weeks post training.

Differences in teaching methods and quality were named in these reviews as factors likely causing the high variation in findings on retention (Vlasblom et al., 2020; Yang et al., 2012). Therefore, the secondary aim of our study was to investigate the influence of two different methods of initial training. In 2020, the CLS initial training in the Dutch Army was reorganized from a 4-week, more classical and classroom-style (frontal) course to a 3-week, more individualized and personalized course. This change was made based on the idea that individualized-style training could be more effective and efficient, and the implementation was accelerated due to the COVID-19 distancing regulations. In the classroom-style course, all trainees receive the same classroom lessons, demonstrations, and exercises, and all trainees follow the same roster. In contrast, the individualized-style group starts out in a classical manner, but is later instead tasked with reflecting upon their progress and needs, and determining with help from the instructor which parts of the material they should spend their time on. This is done with the help of a video application (GPAL, Groningen, The Netherlands), which links instructional videos to the textbook, and allows trainees to record themselves as they perform exercises, and as they present their reflections upon their learning progress and needs. They upload these videos for the instructor to watch, so that personalized feedback and exercises are provided. Trainees also practice together with a buddy in organized practice sessions, where an instructor is present to help and answer questions. The individualized-style course requires more self-direction, responsibility, and ability to reflect on their learning progress from the trainees, while it requires a more coaching and supporting role from the instructors.

Little empirical evidence is available on the effectiveness of these training styles on skills and populations comparable to CLS. Both the individualized-style training and the classroom-style training may have advantages and disadvantages. According to Kirschner et al. (2006), classroom-style "direct instruction" is the most effective way to teach students new skills. Especially young military personnel may prefer and benefit from clear-cut instructions and tasks. However, other authors have shown that, depending on the context of what is being learned, the retention of direct instruction can be low and that such training transfers poorly to new situations (Dean & Kuhn, 2007). Reflecting on performance at the start of training was found to lead to better retention in basic life support skills than individualized feedback during training in medical students by Li et al. (2013). Hamilton (2005) lists several studies on video self-instruction, which suggested that this may lead to better retention of CPR skills than instructor-led classes. If the individualized-style training is organized well, then self-direction and reflection may possibly activate and enhance CLSers' metacognition, play on their individual strengths, and engage them more than classroom-style training (Kuhn, 2007; Vansteenkiste et al., 2006).

Thus, the primary goal of the current study was to add to the existing literature on skill retention in high-risk professions by investigating retention of CLS skills. This part is somewhat explorative, as the wide variety in skill decay outcomes in other studies makes it complicated to predict skill decay for this study. The secondary goal was to investigate whether there is a difference in skill retention between the classroom-style training and the individualized-style training. We predict that the performance immediately after initial training, as well as the 2-month and 6-month retention, is better in the individualized-style training group than in the classroom-style training group based on the literature described above. We therefore expect higher performance quality and speed, and lower perceived mental effort and anxiety during all tests following individualized-style initial training.

Method

Design

A repeated-measures design was used with two participant groups (classroom and individualized). Each participant was tested at three moments, namely at 0 month (i.e., immediately after finishing the course), at 2 months and at 6 months after finishing the course. These intervals were chosen to obtain information on the skill decay curve, as several medical studies found that skill decay did not occur linearly over time, and that most of the decay occurred already within the first 3 months (Vlasblom et al., 2020).

Table 1. Characteristics of the groups.

	Group	
	Classroom	Individualized
Age in years (mean, SD)	22.3 (2.3)	22.7 (2.9)
Work experience in months (mean, SD)	35.4 (25.2)	28.5 (24.0)
Gender (men/women)	22/1	14/3
Education: (secondary/MBO)	4/19	2/15

Participants

The total sample consisted of 40 participants (36 men, 4 women), mean age = 22.5 years, *SD* = 2.52. The highest level of education varied between secondary (17.4%) and vocational education (i.e., "Middelbaar Beroeps Onderwijs" [MBO], 82.6%). Twenty-three participants took part in the classroom-style course and 17 took part in the individualized-style course. All participants had a rank between Private and Private first class. The commissioning of participants over the groups was based on the moment at which the participant started the CLS course during the fall of 2020. As the training style switched from classroom to individualized, later applicants were automatically appointed to the individualized group. The characteristics of each group are shown in Table 1. None of the characteristics differed significantly between the groups as tested with *t*-tests or Chi-squared tests. The study was approved by the ethics committee of TNO, The Netherlands, and all participants provided informed consent before participating.

Apparatus

For testing, we used a Trauma Hal© S3040.100 medical mannequin (Gaumard, Miami, US), dressed in a military uniform that had Velcro strips which allowed for "cutting" the uniform open (see Figure 1). Functions used for the experiment were radial and carotid pulse, chest expansion to simulate breathing and pneumothorax, blinking and closing of the eyes, and pupil responses to light. Changes in vital signs and closing of the eyes were controlled by the instructor using a tablet. The instructor also simulated speech and abnormal breathing sounds of the patient. Other features used were shot- and stab wounds and blast-amputations. Injuries were simulated with fake blood (on the patient's clothing and on the ground) and grime, which included loose skin and dark charring for burn injuries.

A camera (Hero8, GoPro, San Mateo, California, US) filmed the participant's performance from a top-down perspective, for inter-rater reliability testing.

All participants received the same backpack with standard CLS medical equipment. This included chest seals, bandages, burn tecs, gauzes, tapes, disinfectant

Figure 1. A participant performing scenario two on the medical mannequin. Two chest seals are placed on the mannequin's chest. Fake blood from the amputation can be seen on the mannequin's right side, and the medical bag with equipment can be seen at the bottom of the picture.

wipes, nasopharyngeal airways, oropharyngeal airways, decompression needle kit, scissors, a medical flashlight, and pens.

Procedure

The participants performed three tests: at 0 month, 2 months, and 6 months following the end of their initial training. For each test, the participants performed two scenarios of 20 minutes each on the medical mannequin (see Figure 2). Each test consisted of the following chronological parts:

(1) Information was obtained regarding participant's personal characteristics and experience (first test only).
(2) Participants received a briefing with the following instructions. Limitations of the medical mannequin were explained. Participants were told that they could ask the experiment leader to assist with holding something or lifting the mannequin. Participants also received the opportunity to check and rearrange the contents of the backpack with medical supplies.
(3) Participants read the context description (see Table 2), entered the testing room where the

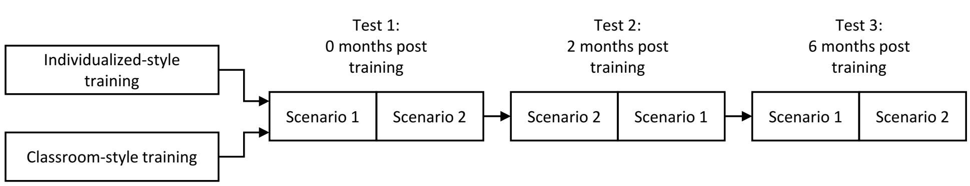

Figure 2. An overview of the full protocol of the study.

Table 2. Context description, as read by the participant before the scenario started. Each description ended with "A helicopter will arrive in 20 minutes and they will help you with transport."

Scenario	Variation	Context description
1	1	"During a foot patrol in a village in Afghanistan, your squad suddenly receives incoming fire. After an intense firefight, the enemy is eliminated. There appears to be someone wounded, and you are ordered to help him."
1	2	"While clearing buildings, you suddenly hear shouting and shots fired. When you arrive at the scene, you see that someone of your squad has been stabbed with a long knife. The enemy is eliminated."
1	3	"During a foot patrol, there is fire contact with the enemy. After the enemy is eliminated, there appears to be somebody wounded by a hand granate. Your commander orders you to treat the wounded."
2	1	"While on a foot patrol, an IED explodes. You rush over and find a colleague confused and injured on the ground. The situation is under control."
2	2	"While driving in a convoy, an IED explodes at the front vehicle. You go to the damaged vehicle and find a colleague confused and injured on the ground. The situation is under control."
2	3	"While clearing buildings there is suddenly an explosion. A confused colleague walks out of the respective building and he falls on the ground. A boobytrap was triggered while searching the rooms. The situation is under control."

mannequin and instructor were, and performed the CLS protocol in the first scenario.
(4) After the scenario finished, participants filled in questionnaires on mental effort and anxiety in a different room (see Dependent measures). They were allowed a break of 10 minutes.
(5) Step 3 was repeated for the second scenario.
(6) Participants were reminded to not share information about the scenarios, and if possible, to not read about or practice the CLS protocol until the last moment of testing. At the end of the last test, they filled in a posttest questionnaire and received feedback from the instructor.

The visit for the first test lasted on average 75 minutes, the second lasted 60 minutes and the third lasted 90 minutes.

Scenarios

Two scenarios were created with the help of CLS instructors, to cover several different injuries and required actions that are highly relevant for CLS. To make the scenarios and events as unpredictable as possible in order to capture realistic performance, the order of the scenarios, the context descriptions (see Table 2) and the source and location of the injuries were changed for each test. The required actions always remained the same.

Scenario 1 consisted of a non-catastrophic bullet or knife wound through the chest, into the front and out the back. There was a non-catastrophic second wound, either on the leg or arm. The patient was able to describe what happened. The patient was unable to breathe deeply or count to 10, had a detectable radial pulse, and a pneumothorax developed when the participant started writing down the life signs (at step E). Radial pulse was then 102 BPM, breathing rate: 45/minute, and Disability classification was "Alert" with ++ pupil response. The patient had no neck pain. The pain score, when asked, was 7/10 and the patient reported pain in the chest (first) and at the location of the second injury if the participant asked for other locations of pain.

Scenario 2 consisted of a blast injury with catastrophic amputation of an underarm or lower leg. There was a burn injury on the face, neck, and mouth with an audible "wheeze" in the breathing. There was a small superficial injury on the chest and two on the limbs. The patient was able to describe what happened. The pain score, when asked, was 8/10, and the patient reported pain at the amputation and, if asked next, on the face. Breathing deeply or counting to 10 was possible. When the participant started checking the Disability classification, the patient lost consciousness and was in shock with no detectable radial pulse. The carotid pulse was then 110 BPM, and the breathing rate was 28/minute.

Dependent measures

Performance quality

To evaluate the performance level of the participants during the tests, their actions were assessed by an instructor using a score form (one for each scenario, see Appendix A). The score form was developed for this study by the authors based on the opinions of CLS instructors while taking the limitations of the used medical mannequin into account. It consists of a number of items for which the instructor selected "yes" or "no" for the participant performing a certain action, or to whether more serious or less serious errors were made in the execution of certain actions. Each item on this initial list was either categorized into critical (i.e., errors cause a direct risk of dying), non-critical (i.e., errors cause an indirect risk of dying or cause a risk of disability) or irrelevant (i.e., removed from the list) in a group discussion with four military-medicine subject-matter experts. The approach to create two categories of items to be analyzed separately was done to avoid having to assign an arbitrary relative weight to the items, which would be needed in order to add them together. The items were re-coded and summed separately for critical and non-critical actions (both scenarios together) to obtain a percentage indicating the proportion of items that were scored positively. A 100% score on both critical and non-critical actions would mean that participants performed the scenario perfectly and completed a second round of writing down vital signs within the available 20 minutes. The performance quality could not be rated by the same instructor during the 2-month test due to unforeseen personal circumstances, and insufficient resources were available to allow for letting this instructor rate the participants using the video footage. Performance quality at 2 months was therefore rated by a different instructor. Inter-rater reliability was determined by comparing the second instructor's scores on each item for nine participants in the 2-month test with those of the first instructor who scored the same participants using the video recordings obtained with the cameras. Cohen's kappa of 0.61 for critical actions and 0.63 for non-critical actions. These values are considered fair to good (Fleiss, 1981).

Performance speed

The instructor noted the time in minutes until the participant reached the Environment/Exposure step. This time frame was chosen because steps <C> ABCD are highly time-critical, whereas E includes continuous monitoring of the patient's vital signs until transport arrives. Since every participant only yields two values of speed per test, we were unable to obtain enough data to perform the inter-rater reliability test for speed.

Mental effort

After each scenario, participants rated how much mental effort they had spent during the scenario on the Rating Scale for Mental Effort (RSME; Zijlstra & Van Doorn, 1985). This is a 150 mm long unidimensional scale ranging from 0 to 150, with nine anchor points ranging from "absolutely no effort" to "extreme effort." The average of the two scenarios was taken to obtain one score for each test.

Anxiety

After each scenario, participants rated how much anxiety they experienced on a 10 cm long horizontal analogue scale with "none" (0) and "maximum" (10) at the endpoints (Houtman & Bakker, 1989). The average of the two scenarios was taken to obtain one score for each test.

Manipulation checks

Several data were obtained to provide info on the extent to which we were able to perform the experiment in a controlled manner. Because it was not possible to schedule each participant exactly at 0, 2, and 6 months, the deviation from these times in weeks was compared between the groups. After the third test, participants were asked if they had either read or practiced the material in the 6 months the experiment ran, and if so, how many times. We also asked participants if they had any foreknowledge about the scenarios, for instance, due to suspecting what would happen based on previous tests or due to talking with colleagues.

Statistical analysis

Performance quality was analyzed using a 3 × 2 × 2 mixed-model ANOVA with two within-subject factors: Test (0, 2, 6 months) and Category (critical, non-critical actions) and one between-subject factor: Group (classroom, individualized). For analyzing performance speed, mental effort and anxiety, a 3 × 2 mixed-model ANOVA was used with one within-subject factor: Test (0, 2, 6 months), and one between-subject factor: Group (classroom, individualized). Mauchly's test of sphericity was used to check sphericity of the data, and Huynh-Feldt adjustment was applied if the data did not meet this criterion. Significant effects were followed-up by post-hoc tests with Bonferroni correction for multiple comparisons. An effect size of $\eta_p^2 = 0.01$ can be considered small, 0.06 as medium, and >0.14 as large (Cohen, 2013).

Results

Performance quality

The Test (0, 2, 6 months) × Group (classroom, individualized) × Category (critical, non-critical) ANOVA revealed a significant main effect of Test, $F(2,76) = 50.27$, $p < 0.001$, $\eta^2_p = 0.57$, a significant main effect of Category, $F(1,38) = 259.35$, $p < 0.001$, $\eta^2_p = 0.87$, a significant interaction effect of Test × Category, $F(2,76) = 6.98$, $p = 0.002$, $\eta^2_p = 0.16$, and a significant three-way interaction effect of Test × Group × Category, $F(2,76) = 8.97$, $p < 0.001$, $\eta^2_p = 0.19$. The in effect of Test, as well as the Test × Category interaction effect, was both overruled by the 3-way interaction effect. The outcomes are shown in Figure 3 and Table 3.

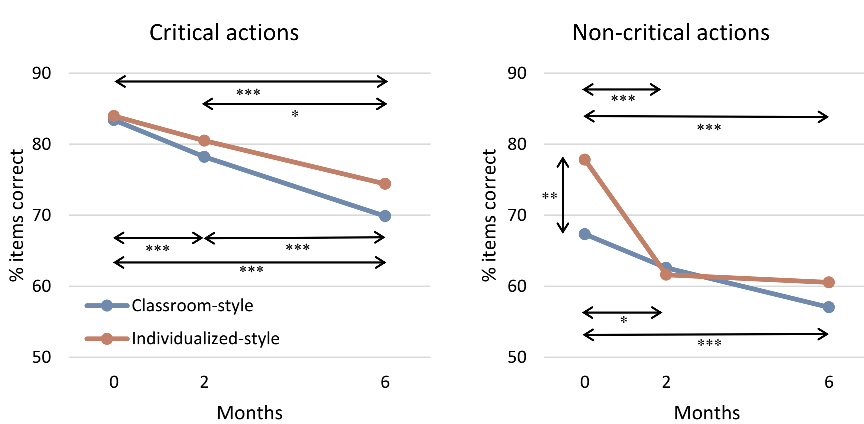

Figure 3. Means of the performance scores for the groups at each test, separated per item category. *p < 0.05, **p < 0.01, ***p < 0.001.

Table 3. Means and SDs of the performance scores (% items correct).

Action category	Group	Test (months)	Mean score (%)	*SD*
Critical	All	0	83.6	8.0
		2	79.4	9.4
		6	72.0	11.4
	Classroom	0	83.4	7.2
		2	78.2	8.2
		6	69.9	10.4
	Individualized	0	84.0	9.2
		2	80.5	10.9
		6	74.4	12.5
Non-critical	All	0	71.8	11.3
		2	62.2	11.3
		6	58.7	10.7
	Classroom	0	67.3	10.4
		2	62.6	10.9
		6	57.0	8.7
	Individualized	0	77.8	9.9
		2	61.6	11.8
		6	60.6	12.9

Group comparisons

Post-hoc comparisons on the 3-way interaction effect showed that there was only a significant difference in that the individualized group scored higher than the classroom group on non-critical actions at 0 month, $\Delta = 10.5\%$, $SE = 3.25$, $p = 0.003$ (see Figure 3). The groups did not score significantly different at any moment of testing on the critical actions.

Test comparisons

Post-hoc comparisons on the 3-way interaction effect revealed that the classroom group had a significant decrease on critical actions score between 0 and 2 months, $\Delta = 5.2\%$, $SE = 1.62$, $p = 0.008$, between 2 and 6 months, $\Delta = 8.4\%$, $SE = 2.22$, $p = 0.002$, and between 0 and 6 months, $\Delta = 13.5\%$, $SE = 1.80$, $p < 0.001$. The individualized group only showed a significant decrease between 0 and 6 months, $\Delta = 9.6\%$, $SE = 2.10$, $p < 0.001$, and between 2 and 6 months, $\Delta = 6.4$, $SE = 2.70$, $p = 0.021$, but not between 0 and 2 months, $p = 0.107$ (see Figure 3).

On non-critical actions, the classroom group's scores decreased significantly between month 0 and 2, $\Delta = 4.7\%$, $SE = 1.71$, $p = 0.026$, and between month 0 and 6, $\Delta = 10.3\%$, $SE = 2.21$, $p < 0.001$, but not between month 2 and 6, $p = 0.052$. This was also the case for the individualized group, $\Delta = 16.2\%$, $SE = 1.99$, $p < 0.001$ (month 0–2), and $\Delta = 17.3\%$, $SE = 2.57$, $p < 0.001$ (month 0–6), $p = 0.545$ (month 2–6).

Item category comparisons

The main effect of Category was not overruled by the interaction effects. Both groups scored significantly higher on critical actions than on non-critical actions at each time of measurement, $\Delta = 13.9\%$, $SE = 0.86$, $p < 0.001$.

Performance speed

The Test (0, 2, 6 months) × Group (classroom-style, individualized-style) repeated-measures ANOVA revealed a significant main effect of Test, $F(2,66) = 5.80$, $p = 0.005$, $\eta^2_p = 0.15$, a significant main effect of Group, $F(1,33) = 5.14$, $p = 0.030$, and no significant Test × Group interaction effect, $F(2,66) = 1.06$, $p = 0.352$. Post-hoc tests on Test revealed that there was a significant increase in performance speed at 2 months compared to 0 month, $\Delta = 1.2$ minutes, $SE = 0.30$, $p = 0.001$, and no other significant differences. The post-hoc test on Group revealed that the individualized-style group was significantly faster than the classroom-style group, $\Delta = 1.2$ minutes, $SE = 0.52$, $p = 0.030$. The outcomes are shown in Figure 4.

Mental effort

The Test (0, 2, 6 months) × Group (classroom, individualized) repeated-measures ANOVA on mental effort ratings revealed a significant main effect of Test, $F(2,76) = 3.64$, $p = 0.031$, no significant main effect of Group, $F(1,38) = 0.40$, $p = 0.530$, and a significant Test × Group interaction effect, $F(2,76) = 4.12$, $p = 0.020$, which overruled the significant main effect of Test. Post-hoc comparisons showed that there was a significant increase in mental effort in the classroom group from month 0 to 6, $\Delta = 18.8$, $SE = 4.9$, $p = 0.001$, while there were no significant differences for the individualized group. At none of the tests, there was a significant difference between the groups (see, Figure 5).

Anxiety

The Test (0, 2, 6 months) × Group (classroom, individualized) repeated-measures ANOVA on

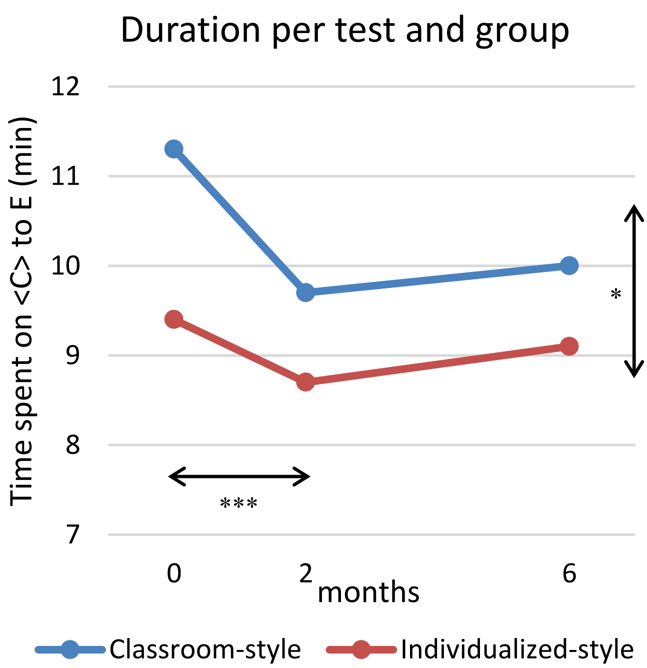

Figure 4. Mean duration of performing the first part of the protocol (<C> to E) at each test. *p < 0.05, **p < 0.01, ***p < 0.001.

anxiety ratings revealed a significant main effect of Test, $F(2,76) = 4.40$, $p = 0.016$, no significant main effect of Group, $F(1,34) = 0.01$, $p = 0.907$, and a significant Test × Group interaction effect, $F(2,68) = 3.55$, $p = 0.034$ which overruled the significant main effect of Test. Post-hoc comparisons revealed that the classroom-style group reported a significant increase in anxiety from 0 to 2 months, $\Delta = 1.45$, $p = 0.014$, and from 0 to 6 months, $\Delta = 1.57$, $p = 0.001$, while there were no significant differences for the individualized group (see, Figure 5). At none of the tests, there was a significant difference between the groups.

Manipulation checks

Practice during the experiment

In both groups, the median number of times participants had read or practiced the material in the 6 months during the experiment was 2. A Mann-Whitney *U* test revealed no significant difference between the groups, $U = 172.00$, $p = 0.846$.

Foreknowledge about the scenarios

Overall, little foreknowledge was reported and reported foreknowledge was slightly higher than the classroom group. None of the individualized group reported having had foreknowledge about the scenarios at the tests. In the classroom group, one participant reported having had "medium amount of foreknowledge" at the 0-month test, three participants reported having had "little foreknowledge" at the 2-month test and two reported having had "little foreknowledge" at the 6-month tests.

Test moments

The exact test moments in weeks, as well as a comparison between groups at each test moments (Mann-Whitney U-test), are shown in Table 4. The comparisons showed that the individualized group performed the 2-month test significantly later than the classroom test by approximately 1 week. However, the performance decrease between 0 month and 2 months was not significantly correlated with duration between the tests in the classroom group, p's > 0.638, and in the individualized group, p's > 0.549.

Missing data

Five cases of missing outcomes of a test were imputed for performance quality, six for performance speed, and four for mental effort and anxiety. Four of these cases were due to participants being unavailable for the

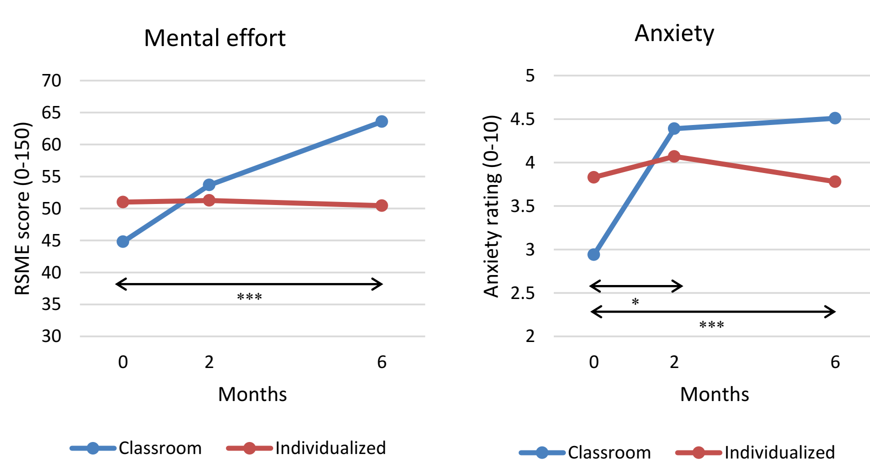

Figure 5. The mean mental effort and anxiety ratings for the three groups at each test. *p < 0.05, **p < 0.01, ***p < 0.001.

Table 4. The test moments in weeks since finishing the course. Ideally, the 0-month test should ideally be at 0 week, the 2-month test at 8.69 weeks, and the 6-month test at 26.07 weeks.

	Group							
	Classroom			Individualized				
	Mean (SD)	Median	*N*	Mean (SD)	Median	*N*	*U*	*p*
0-month test (weeks)	0.20 (0.89)	0.00	23	1.13 (1.89)	0.00	17	155.00	0.094
2-month test (weeks)	8.65 (1.73)	8.00	21	9.25 (1.39)	9.00	16	100.50	0.021
6-month test (weeks)	27.00 (4.82)	25.00	21	27.31 (3.11)	26.00	17	140.50	0.259

concerning test due to quarantining for Covid or other obligations. Other missing data were caused by measurement errors. Imputations were performed by replacing missing values by means of the group at that test.

Discussion

The primary goal of the current study was to measure the retention of CLS skills over a period of 6 months. The results of the experiment revealed that performance of critical actions had decreased by 5% at 2 months and by 14% at 6 months compared to performance at 0 month after initial training (Table 3). The average number of critical errors per scenario increased from 3.4 at 0 month to 5.8 at 6 months, a number which could have serious consequences in operational practice. For non-critical actions, the decline was steeper in the first 2 months, as performance at 2 months had decreased by 13% and at 6 months by 18%.

The secondary goal of the study was to test if retention would be higher in a more individualized-style training group compared to a more classroom-style trained group. There were few differences between the groups in favor of the individualized group. On non-critical actions, the individualized group performed significantly better than the classroom group at 0 month only, after which their performance dropped to a similar level. This suggests that individualized training leads to higher initial performance, but that this higher level of performance was not maintained. The individualized group performed the protocol significantly faster overall than the classroom group, while they did not make significantly more critical or non-critical errors. This implies that their performance was more efficient. The classroom group also reported a gradual increase in mental effort and anxiety for each subsequent test, whereas the individualized group did not. This suggests that the tests became more demanding over time for the classroom group, although differences between groups did not reach significance.

The decline in performance in the examined time interval, as well as the large variance in performance at each moment of testing, are in line with other findings with regard to complex skills (Arthur et al., 1998; Sanli & Carnahan, 2018). In contrast to several other studies in safety-critical professions (see Vlasblom et al., 2020), we did not find evidence of the skill decay curve reaching a steady level before 6 months for the critical actions, although performance on non-critical actions declined more steeply in the first 2 months than in the next 4 months. Whereas skill decay in other high-risk professions was found to be around 30–60% at 6 months (Vlasblom et al., 2020; Figure 2(b) on p 8), skill decay in our population was lower (i.e., 14–18% of initial skill level, see Table 3). This could indicate that our population was relatively well trained, although this discrepancy may also result from our manner of performance scoring, which only took into account those actions that either directly (i.e., critical actions) or indirectly (i.e., non-critical actions) endangered the patient's life. The high number of critical errors at 6 months in our population (i.e., 5.8 on average) is still alarming.

Based on this and based on the finding that the advantage of the individualized-style training dropped sharply in the first 2 months, we advise that extra refresher training is organized within 2 months following the initial training. The 3–4 weeks available for initial training may not allow for enough time-interval or "spacing" between exercises. Spacing was previously shown to positively impact retention of medical skills, such as surgery (Cecilio-Fernandes et al., 2018), and some studies have found benefits of increased spacing for up to 60 days after the last refresher training (see Roediger, Nestojko, & Smith, 2019). Extra early refresher training organized in the CLSers' own units could possibly combat this issue and improve retention further down the line. When organizing refresher training within the CLSers' own units, the training could possibly be integrated within other military exercises, so that the skills are practiced in a realistic context. Other more low-cost options to provide refresher training involve the use of instruction videos and serious games. Four of such 30-minute refresher sessions in the 2 weeks following initial training were found to have a positive effect on retention of more basic CLS skills that are taught to all military personnel (Planchon et al., 2018).

There were several other interesting observations with regard to the outcomes. First, the initial performance of our CLS sample was already low immediately after initial training, as the full sample made on average 3.4 critical errors per scenario at 0 month. It seems that the initial training in our sample was insufficient to prevent multiple critical errors immediately following the training, and it seems also that the individualized-style training was no more effective in this regard. The effectiveness of the individualized-style training could possibly be improved by supporting the instructors better in organizing this style of training, which was new to them. This could also improve the instructors' attitude toward the change, which was negative for some based on anecdotal data. Possibly, the training needs to evolve more to create a better balance between individualized and classroom sessions. Categorizing actions into critical and non-critical can possibly help to create consensus among instructors and to provide clarity to students.

A second observation is that performance on non-critical actions was generally lower than that on critical actions. This can be explained by the fact that performance was measured up until the completion of one second round of checking the vital signs, and many participants did not complete this within the allotted time.

A third observation is that performance speed in both groups increased in the 2-month test compared to the 0-month test and did not change significantly from month 2 to 6. This could indicate that more items were being skipped at 2 months. This could not be confirmed by performance data as non-adherence to an item can mean that an erroneous attempt was made to perform the action or that the action was skipped completely. It is, however, important to note that speed data at the 2-month test was obtained from a second instructor, and no inter-rater reliability check on the speed was performed due to insufficient resources.

The current study is limited by the use of subjective performance measures. A second instructor rated the 2-month test, and correlation between scoring of individual items was not high, possibly due to differences between instructors in scoring actions that were performed at a later time then required according to the protocol. This should be taken into account when interpreting the results of the 2-month test. A second limitation is that we did not succeed in completely preventing the participants from practicing or reading the material during the 6 months of the experiment. The number of such reported instances was very low (i.e., median = 2), but this could have slightly affected the results.

In conclusion, based on our findings with regard to CLSers' skill decay over the course of 6 months, we advise to organize extra refresher training immediately following the initial training, and to further develop and optimize the individualized-style training, which had some positive effects.

Acknowledgments

We want to express our special thanks to Renee van Deursen (Radboud University, Faculty of Social Sciences, Nijmegen, The Netherlands) for her excessive support during the preparation and execution of the experiment. We also want to gratefully thank Major Willem Goes (Dutch Ministry of Defense Health Organization Training Center, Hilversum, The Netherlands) for his excellent work as point of contact between the Training Center, 13th infantry brigade and the research team. We appreciatively thank Marco Politowski (Triangular Group Academy, Heerle, The Netherlands) for his assistance as medical instructor and main observer, and we like to give our special thanks to Dr Esther Oprins (TNO Soesterberg, The Netherlands) for providing indispensable advice and her trust in this research project.

Data availability statement

Raw data (video recordings) cannot be shared to protect the privacy of the participants.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by the Dutch Ministry of Defense, under program: V1806: Education & Innovative Training.

ORCID

Annemarie Landman http://orcid.org/0000-0003-3678-9210
Olaf Binsch http://orcid.org/0000-0002-9395-9741

References

Arthur, W., Jr, Bennett, W., Jr, Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention:

A quantitative review and analysis. *Human Performance*, *11* (1), 57–101. https://doi.org/10.1207/s15327043hup1101_3

Cecilio-Fernandes, D., Cnossen, F., Jaarsma, D. A., & Tio, R. A. (2018). Avoiding surgical skill decay: A systematic review on the spacing of training sessions. *Journal of Surgical Education*, *75*(2), 471–480. https://doi.org/10.1016/j.jsurg.2017.08.002

Center for Army Lessons Learned. (2017). Tactical Combat casualty care handbook: Lessons and best practices (Version 5, No, 17-13). https://usacac.army.mil/sites/default/files/publications/17493.pdf

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

Dean, D., Jr, & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education*, *91*(3), 384–397. https://doi.org/10.1002/sce.20194

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). John Wiley.

Hamilton R. (2005). Nurses' knowledge and skill retention following cardiopulmonary resuscitation training: a review of the literature. J Adv Nurs, 51(3), 288–297. https://doi.org/10.1111/j.1365-2648.2005.03491.x

Hancock, P. A., & Szalma, J. L. (2008). *Performance under stress*. Ashgate publishers.

Houtman, I. L. D., & Bakker, F. C. (1989). The anxiety thermometer: A validation study. *Journal of Personality Assessment*, *53* (3), 575–582. https://doi.org/10.1207/s15327752jpa5303_14

Kirschner, P., Sweller, J., & Clark, R. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Kuhn, D. (2007). Is direct instruction an answer to the right question?. *Educational Psychologist*, *42*(2), 109–113. https://doi.org/10.1080/00461520701263376

Landman, A., van Middelaar, S. H., Groen, E. L., van Paassen, M. M., Bronkhorst, A. W., & Mulder, M. (2020). The effectiveness of a mnemonic-type startle and Surprise management procedure for pilots. *The International Journal of Aerospace Psychology*, *30*(3–4), 104–118. https://doi.org/10.1080/24721840.2020.1763798

Linde, A. S., Caridha, J., & Kunkler, K. J. (2018). Skills decay in military medical training: A meta-synthesis of research outcomes. *Military Medicine*, *183*(1–2), e40–e44. https://doi.org/10.1093/milmed/usx023

Li, Q., Zhou, R. H., Liu, J., Lin, J., Ma, E. L., Liang, P., . . . Xiao, H. (2013). Pre-training evaluation and feedback improved skills retention of basic life support in medical students. *Resuscitation*, *84*(9), 1274–1278. https://doi.org/10.1016/j.resuscitation.2013.04.017

Oprins, E., Mol, D., & Bottenheft, C. (2019). Competence retention in military helicopter underwater egress training (HUET). Abstract of 61th International Military Testing Association Conference, 7-11 October Tallin, Estonia.

Planchon, J., Vacher, A., Comblet, J., Rabatel, E., Darses, F., Mignon, A., & Pasquier, P. (2018). Serious game training improves performance in combat life-saving interventions. *Injury*, *49*(1), 86–92. https://doi.org/10.1016/j.injury.2017.10.025

Roediger, H. L., Nestojko, J. F., & Smith, N. S. (2019). Strategies to improve learning and retention during training. *Human Performance Optimization*, 302–332.

Sanli, E. A., & Carnahan, H. (2018). Long-term retention of skills in multi-day training contexts: A review of the literature. *International Journal of Industrial Ergonomics*, *66*, 10–17. https://doi.org/10.1016/j.ergon.2018.02.001

Sullivan, A., Elshenawy, S., Ades, A., & Sawyer, T. (2019). Acquiring and maintaining technical skills using simulation: Initial, maintenance, booster, and refresher training. *Cureus*, *11*(9). https://doi.org/10.7759/cureus.5729

Vansteenkiste, M., Lens, W., & Deci, E. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Educational Psychologist*, *41*(1), 19–31. https://doi.org/10.1207/s15326985ep4101_4

Vlasblom, J. I., Pennings, H. J., van der Pal, J., & Oprins, E. A. (2020). Competence retention in safety-critical professions: A systematic literature review. *Educational Research Review*, *100330*. https://doi.org/10.1016/j.edurev.2020.100330

Wisher, R. A., Sabol, M. A., & Ellis, J. A. (1999). Staying sharp: Retention of military knowledge and skills. ARI Special Report no. 39. U.S. Army Research Institute.

Yang C, Yen Z, McGowan J E, Chen H Carrie, Chiang W, Mancini M E, Soar J, Lai M and Ma M Huei-Ming. (2012). A systematic review of retention of adult advanced life support knowledge and skills in healthcare providers. Resuscitation, 83(9), 1055–1060. https://doi.org/10.1016/j.resuscitation.2012.02.027

Zijlstra, F. R. H., & Van Doorn, L. (1985). The construction of a scale to measure perceived effort. Technical Report. Delft University of Technology.

Appendix A. Items on the scoring form

Category 1 is critical, category 2 is non-critical.

*Item score was reversed.

Scenario 1: Chest wound and pneumothorax

Letter of the protocol	Item	Category
<C>	Starts with complete "status, report, triage"	2
<C>	Checks first for catastrophic hemorrhage	1
<C>	Diagnoses catastrophic hemorrhage incorrectly*	2
<C>	Treats presumed catastrophic hemorrhage according to protocol	2
A	Asks: "What happened?"	2
A	Asks: "Where are you hurting?"	2
A	Asks: "Where else are you hurting?"	2
A	Asks: "Is your neck hurting?"	2
A	Unnecessarily fixates the neck*	2
A	Listens for abnormal sounds in patient breathing	1
A	Reports: no extra sounds in breathing	2
A	Inspects inside of the mouth	1
A	Instructs: "If there is something in your mouth, please spit it out."	1
B	Asks to sigh deeply or count to ten	1
B	Decides to perform chest inspection	1
B	Visually checks for open pneumothorax	1
B	Makes serious error in placing chest seal, or does not place chest seal on chest wound*	1
B	Places chest seal on patient back	1
B	Manually checks for pneumothorax	1
B	Manually checks for flail chest	2
C	Checking for other serious hemorrhage is above average	2
C	Checking for other serious hemorrhage is below average*	1
C	Treats other serious hemorrhage	2
C	Checks radial pulse	1
C	Incorrectly reports shock	2
D	Checks alertness with question(s)	2
D	Makes correct disability classification (Alert)	1
D	Checks pupil response	1
D	Reports: "T1 patient with B and C problem."	1
D	Asks for transport capacity	2
E	Checks for burn injury	2
E	Protects patient from environment	1
E	Makes serious error when checking vital signs or does not check vital signs*	1
E	Makes small error when checking vital signs	2
B	Detects pneumothorax	1
B	Makes a serious error in placement of ARS needle, of does not place needle*	1
E	Asks for pain score	2
E	Administers correct pain medication	2
E	Performs body check (for fractures etc.)	2
E	Makes serious error in second check of vital signs, or does not perform second check of vital signs*	2
E	Makes small error in second check of vital signs or does not perform second check of vital signs*	2
E	Fills in data on patient card	2
E	Makes serious error in transfer report*	2
E	Makes small error in transfer report*	2

Scenario 2: Amputation, burn injury, and loss of consciousness.

Letter of the protocol	Item	Category
<C>	Starts with complete "status, report, triage"	2
<C>	Checks first for catastrophic hemorrhage	1
<C>	Makes a serious error in placing the tourniquet, or does not place tourniquet*	1
<C>	Makes a small error in placing the tourniquet, or does not place tourniquet*	2
A	Asks: "What happened?"	2
A	Asks: "Where are you hurting?"	2
A	Asks: "Where else are you hurting?"	2
A	Asks: "Is your neck hurting?"	2
A	Unnecessarily fixates the neck*	2
A	Listens for sounds in patient breathing	1
A	Identifies inhalation trauma	2
A	Reports inhalation trauma to commander	1
A	Inspects inside of the mouth	1
A	Instructs: "If there is something in your mouth, please spit it out."	1
B	Asks to sigh deeply or count to ten	1
B	Unnecessarily performs chest inspection*	2
C	Checks tourniquet	1
C	Checking for other serious hemorrhage is above average	2
C	Checking for other serious hemorrhage is below average*	1
C	Treats other serious hemorrhage	2
C	Checks radial pulse	1
D	Checks alertness with question(s)	2
D	Returns in protocol due to loss of consciousness	1
<C>	Checks tourniquet	1
A	Makes serous error in placing of NPA or OPA, or does not place NPA or OPA*	1
B	Checks vital signs for 15 seconds*	1
C	Reports shock	2
D	Checks disability	1
D	Makes correct disability classification (Alert)	2
D	Administers pain stimulus	2
D	Checks pupil response	1
D	Reports: "T1 patient with C and D problem."	1
D	Mentions shock in report	2
D	Asks for transport capacity.	1
E	Checks for burn injury.	1
E	Treats burn injury effectively	1
E	Keeps the time when treating burn injury	2
E	Protects patient from environment	1
E	Makes serious error when checking vital signs or does not check vital signs.*	1
E	Makes small error when checking vital signs.	2
E	Performs body check (for fractures etc.)	2
E	Makes serious error in second check of vital signs, or does not perform second check of vital signs.*	2
E	Fills in data on patient card	2
E	Makes serious error in transfer report	2
E	Makes small error in transfer report	2