



# Is there one best way to support skill retention? Putting practice, testing and symbolic rehearsal to the test

Barbara Frank<sup>1</sup> · Annette Kluge<sup>1</sup>

Published online: 21 December 2018  
© The Author(s) 2018

## Abstract

While a great deal is already known about the effectiveness of training delivery methods, the effectiveness of methods to support skill retention has not yet been sufficiently examined. To address this gap, three studies with different task types were conducted, comprising a total of 240 participants (80 per study). Participants learned how to perform a simulated process control task, which served as a prototype for a setting prone to skill decay. The aim was to compare three refresher interventions (Practice, Testing, and Symbolic Rehearsal), which differ in their underlying theoretical rationale. Participants in all three studies learned a task in week 1 (Study 1: fixed-sequence task, Study 2: contingent-sequence task, Study 3: parallel-sequence task). In each study, participants were divided into four equal-sized groups, which received either no refresher intervention or one of the following three refresher interventions one week after initial training (week 2): Practice, Skill Test, Symbolic Rehearsal. After two weeks, they performed the task again without help (week 3). Independently of the task, refresher interventions reduced the number of mistakes, especially when a Practice refresher intervention was applied. The classical “testing effect” could not be replicated.

*Practical Relevance:* Independently of the task, refresher interventions reduced the number of mistakes, especially when a Practice refresher intervention was applied. The classical “testing effect” could not be replicated.

**Keywords** Refresher training · Testing effect · Practice · Symbolic Rehearsal · Errors in primary task performance

---

Dr. Barbara Frank  
barbara.frank@rub.de

✉ Prof. Dr. Annette Kluge  
annette.kluge@rub.de

<sup>1</sup> Department of Work, Organizational and Business Psychology, Faculty of Psychology, Ruhr-University Bochum, Universitätsstraße 150, 44801 Bochum, Germany

## Gibt es den einen besten Weg für den Fertigkeitserhalt? Üben, Testen und symbolische Wiederholung im Vergleich

### Zusammenfassung

Obwohl bereits vieles über die Wirksamkeit von Trainingsmethoden bekannt ist, wurde die Wirksamkeit von Methoden zur Unterstützung des Kompetenzerhalts noch nicht ausreichend untersucht. Um diese Lücke zu schließen, wurden drei Studien mit unterschiedlichen Aufgabentypen durchgeführt, die insgesamt 240 Teilnehmer (80 pro Studie) umfassten. Die Teilnehmenden lernten, wie man eine simulierte Prozesssteuerungsaufgabe ausführt, die prototypisch für Fertigkeiten ist, die, wenn sie nicht aufgefrischt werden, anfällig für Fertigkeitsverlust sind. Ziel war es, drei Refresher-Interventionen (Üben, Testen und symbolisches Wiederholen) zu vergleichen, die sich in den zugrunde liegenden theoretischen Prozessen unterscheiden. Die Teilnehmenden aller drei Studien lernten in der ersten Woche eine Aufgabe (Studie 1: Aufgabe mit fester Sequenz, Studie 2: Aufgabe mit bedingter Sequenz, Studie 3: Parallelsequenz-Aufgabe). In jeder Studie wurden die Teilnehmenden in vier vergleichbare Gruppen eingeteilt, die entweder eine keine Refresher-Intervention oder eine der folgenden drei Refresher-Interventionen eine Woche nach dem ersten Training (Woche 2) erhielten: Üben, Fertigkeitstest, symbolische Wiederholung. Nach zwei Wochen führten sie die Aufgabe erneut aber ohne Hilfe durch (Woche 3). Unabhängig von der Aufgabe reduzierten Refresher-Interventionen die Anzahl der Fehler, insbesondere bei erneuten Üben. Der „Test-Effekt“ konnte nicht repliziert werden.

*Praktische Relevanz:* Unabhängig von der Aufgabe reduzieren Refresher-Interventionen die Anzahl der Fehler, insbesondere bei weiterer Übung. Der klassische „Test-Effekt“ konnte nicht repliziert werden.

**Schlüsselwörter** Refresher-Training · Test-Effekt · Üben · Symbolische Wiederholung · Fehler in der Hauptaufgabe

## 1 Background

Organizations expect lifelong learning and lifelong remembering from their employees. But many skills are only required infrequently, for instance due to a high level of automation in production (Kluge et al. 2009), demands of worker flexibility and a high task variety (Karuppan 2011), or long periods of non-use during daily operations. Accordingly, skill retention becomes a challenge (Arthur and Day 2013a; Arthur et al. 2010; Kim et al. 2013). In some organizations, knowledge and skill retention are crucial for safeguarding people’s lives, for example in hospital emergency rooms (e. g. Kaye et al. 1987; Perez et al. 2013; Wollard et al. 2006), military and combat environments (e. g. Arthur et al. 1998; Farr 1987; Hoffman et al. 2010; Villado et al. 2013; Wang et al. 2013), general aviation (Casner et al. 2013; Childs et al. 1983; Fanjoy and Keller 2013; King 2015; Walley 1995) and process industries, e. g. chemical or power plants (Bainbridge 1983; Kluge and Frank 2014; Kluge et al. 2014, 2015). Moreover, skill retention also needs to be supported in modern advanced manufacturing technology environments (Jaber and Kher 2004; Karuppan 2011; Nembhard and Uzumeri 2000).

But what is the best way to support skill retention? As the cited literature and previous articles (Kluge and Frank 2014; Kluge et al. 2015) sum up the current state-of-the-art concerning factors influencing skill retention, in the present paper, we conduct a comparative investigation of the design of interventions to support skill retention by way of three studies.

**Use of refresher interventions to prevent loss of access to skills after periods of non-use** According to the New Theory of Disuse, learned knowledge and skills, once acquired, remain in long-term memory, and difficulties in recalling knowledge and skills are entirely determined by the current retrieval strength (Bjork 2011). The problem is not the “loss” of knowledge and skill per se, but rather the loss of access (Bjork 2009, 2011; Bjork and Bjork 1992). Access to items in memory is lost due to interference from competing information and altered stimulus conditions such as recency and current cues (Kluge 2014). As one learns new information, procedures and skills, there is potential for competition with related information, procedures and skills that already exist in memory (Bjork 2011). Arthur and Day (Arthur and Day 2013b) describe decay as a matter of competition between relevant and irrelevant information and processes stored in memory. As new items are learned and added to memory, or as the retrieval strength of certain items is increased, for example by more frequent recall, other items become less retrievable (Bjork and Bjork 1992).

Loss of access can be prevented by interventions to refresh the access and to retain a high retrieval strength, e. g. by means of repetition, retraining and rehearsal (Stammers 1981), also called refresher interventions. An intervention is “each kind of externally controlled, goal-oriented and systematic impact on individuals [...]” that consists of a number of tasks with instructions (Hager and Hasselhorn 2008, p. 41). A refresher intervention (RI) aims to re-establish a specific skill level that was acquired at the end of initial training, which should be re-established after a period

of non-use in which recall of the skill was not required (Kluge et al. 2012, p. 2437).

Research on the methods of delivery of refresher interventions has been neglected so far. While there is a comprehensive body of literature explaining conditions and effects of training and development (Bell et al. 2017), studies which attempt to understand and explain the design and delivery methods of refresher interventions are relatively sparse. Exceptions are papers by Bodilly et al. (1986), Ginzburg and Dar-El (2000), Schendel and Hagman (1982). It is commonly agreed that “the effectiveness of a particular training delivery method depends on the skill or task being trained and the desired learning outcomes” (Bell et al. 2017, p. 315). In contrast, with respect to refresher interventions, organizations seem to show a more or less common sense-based application of delivery methods, while evidence-based decisions regarding delivery seem to be few and far between. The objective of the present paper is to compare three different refresher intervention delivery methods regarding their impact on skill retention. Such a comparison should help practitioners to select refresher intervention delivery methods in an evidence-based manner.

To systematically investigate the effects of refresher intervention methods on skill retention, we selected the three which are assumed to be the most effective (e. g. Farr 1987): 1. Practice, based on the theoretical assumptions regarding skill proceduralisation (Anderson 1982; Schneider 1985; Sun et al. 2001, 2005); 2. Testing, based on the theoretical assumptions of the testing effect (Bjork and Bjork 2006; Roediger and Karpicke 2006); and 3. Symbolic Rehearsal, based on the theoretical assumption of mental practice (Driskell et al. 1994; Farr 1987; Naylor and Briggs 1961).

**The underlying principles of practice, testing and symbolic rehearsal** Practice during initial training means repetitive work on a task until a certain proficiency level is reached (Kluge et al. 2009). A Practice refresher intervention is a practical repetition of learned material (Gonzalez et al. 2003; Kluge 2014), which supports the additional proceduralisation of skills (Anderson 1982; Kim et al. 2013). Repetitions have proven to be effective when applied both before and after task proficiency has been achieved (Hagman and Rose 1983). When used as a refresher intervention, Practice can be regarded as a variant of “distributed practice”, in which Practice is interrupted by longer periods of non-use. In general, retention can be enhanced by increasing the amount of training through task repetition, and studies have provided consistent evidence that Practice refresher interventions support skill retention (e. g. Foss et al. 1989; Kontogiannis and Shepherd 1999; Mattoon 1994; Morris and Rouse 1985; Kluge and Frank 2014; Sauer et al. 2008). Testing skills is proposed as a promising approach to sup-

port knowledge and skill retention (Bjork and Bjork 2006; Farr 1987; Roediger and Karpicke 2006) and is assumed to be superior to Practice. Research results on the “testing effect” (Roediger and Karpicke 2006) from the field of instructional psychology suggest that Tests that are applied after an initial training phase support knowledge retention more strongly than additional practice of the learning material (McDaniel et al. 2007; Pashler et al. 2007; Roediger and Karpicke 2006). The testing effect is explained by 1) the intense retrieval effort which learners have to invest in the testing situation in order to retrieve information from long-term memory, and 2) a transfer-enhancing processing of information, which is identical in the refresher situation and the later retention assessment situation (Bjork and Bjork 2006; Roediger and Karpicke 2006). So far, the testing effect has mainly been investigated using traditional, non-dynamic verbal learning material, and findings on the superiority of Tests over Practice for complex skills are inconsistent (Karpicke and Aue 2015; Leahy et al. 2015; van Gog et al. 2015; van Gog and Sweller 2015). Our own previous research showed that Testing a skill (a Skill Test refresher intervention) by executing the procedure is more effective than Testing the procedural knowledge only theoretically without executing the procedure (Kluge and Frank 2014).

Refresher interventions can also be designed as Symbolic Rehearsal (Driskell et al. 1994; Farr 1987; Naylor and Briggs 1961), in which a person visualizes how to perform a task, takes notes, or makes a drawing of how to perform a task without actually performing it (Annett 1979; Driskell et al. 1994). Symbolic Rehearsal appears to be particularly promising for the retention of cognitive skills, mental operations (Driskell et al. 1994; Farr 1987; Kluge et al. 2012; Naylor and Briggs 1961) and sequences of actions (Cooper et al. 2001). Previous research by Kluge et al. (2012) and Kluge et al. (2015) showed that Symbolic Rehearsal attenuates skill decay, but is less effective than Practice and Testing (Naylor and Briggs 1961).

As the objective of our research was to find the “one best way” for skill retention, we conducted three studies in order to compare these three refresher interventions applied to three different tasks (a fixed, contingent and parallel sequence task) which are e. g. required in process control environments.

The research cited above originates from different learning and retention contexts, such as military skills and basic educational/school contexts. Thus, it is unsurprising that answers to the question of which refresher intervention is the most effective are mixed. The strongest evidence can be found for the hypothesis that refresher interventions in general support skill retention: Refresher interventions counteract skill decay (H1). The evidence for Practice refresher interventions is unanimous and consistent, whereas the em-

**Table 1** Description of tasks  
**Tab. 1** Beschreibung der Aufgaben

Tasks	Description
Fixed-sequence task	The operator first needs to ascertain what kind of task has to be executed (e. g. start-up of a plant or error management) and then needs to execute the initially learned standard operating procedures sequentially (“If S1, then x.”; Omerod and Shephard 2004; Kluge 2014)
Contingent-sequence task	Contingent-sequence tasks can be defined by multiple, interdependent and real-time decisions, occurring in an environment that changes independently and as a function of a sequence of actions (Brehmer 1992) and means “If S2, either z, then x; or not z, then y.” (Omerod and Shephard 2004; Kluge 2014). In such an environment, decisions under certainty take place: The operator is aware of possible alternatives, consequences and the order of preferences (Dörner and Bick 1994). A contingent-sequence task under certainty can consist of a fixed-sequence task in which, at a special point or under a special condition, the operator has to perform the next steps based on a correct gathering of information and interpretation of the situation
Parallel-sequence task	Basically, consist of two sequences which have to be synchronized in time (Proctor and Dutta 1995; Wickens 2008; Wickens and McCarley 2008) and means “If S3, then do x and y” (Omerod and Shephard 2004; Kluge 2014). In these tasks, for example, the operator has to control a second task while executing a first task, and both tasks are executed based on Standard operating procedures. A conscious, directed attention allocation and time-sharing is necessary to perform the task (Schumacher et al. 2001; Wickens and McCarley 2008). An example of such a task is when a pilot is controlling different instruments during take-off, and consequently has to divide his/her attention according to change frequency and how valuable and costly the attention is (Moray 1996)

irical evidence for the testing effect and the superiority of Testing is mixed. Therefore, we assume that: Practice refresher interventions support skill maintenance better than Skill Test refresher interventions (H2). Finally, based in particular on our own research showing that Testing is more effective than Symbolic Rehearsal, we assume that Skill Test refresher interventions support skill maintenance more efficiently than Symbolic Rehearsal refresher interventions (H3).

## 2 Method

Refresher interventions are especially important for procedural skills, which are susceptible to skill decay even after short periods of non-use (Farmer et al. 1999; Farr 1987). In the present studies, the participants were required to learn such a complex skill. Complex skills were applied in a fixed-sequence, a contingent-sequence, and a parallel-sequence task (Omerod and Shephard 2004). The tasks are summarized in Table 1.

Based on findings from preliminary studies, mean effect sizes were expected ( $\eta^2_p$  between 0.13 and 0.45, see Kluge et al. 2012). Therefore for the experiments, according to a mean effect size  $f$  of 0.25, a significance level of alpha of 0.05, a test strength of 0.95 is assumed and from this a total sample size of at least 80 subjects was determined for each study (Faul et al. 2007).

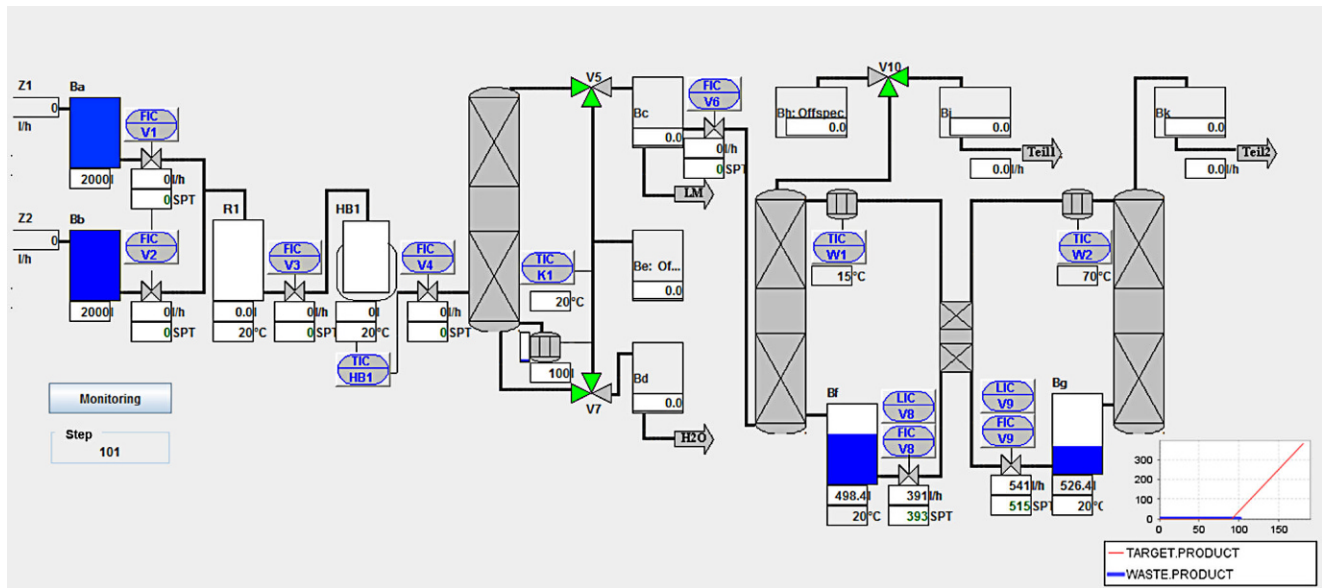
### 2.1 Samples

Studies 1 to 3 were conducted from October 2014 to December 2015. The three studies differed only with respect to the learned task type (Table 2).

Study 1. Eighty engineering students (24 female) from the Ruhr-University Bochum participated in the study from October 2014 to December 2014. The participants were recruited by postings on social networking sites and flyers handed out on the university campus. To ensure basic technical understanding, only engineering students were eligible to participate. Participants received 25 € (Control Groups) or 30 € (Refresher Groups) for taking part. The study was approved by the local ethics committee. Participants were

**Table 2** Samples of studies 1–3  
**Tab. 2** Stichproben der Studien 1–3

	Task	Refresher Intervention	Sample	Design	Selection criterion
Study 1	Fixed-sequence task	Practice Skill Test Symbolic Rehearsal Control	80 engineering students (excl. 3)	Between-within-subject design	Minimum production outcome $\geq$ 2001
Study 2	Contingent-sequence task	Practice Skill Test Symbolic Rehearsal Control	80 engineering students (excl. 4)	Between-within-subject design	Minimum production outcome $\geq$ 1001
Study 3	Parallel-sequence task	Practice Control Symbolic Rehearsal Skill Test	80 engineering students (excl. 7)	Between-within-subject design	Minimum production outcome $\geq$ 2001



**Fig. 1** WaterSim Screenshot with valves (V1–V9), heaters (HB1, K1, W1, W2) and tanks (Ba, Bb, R1, HB1, Bc, Be, Bd, Bh, Bj, Bk, Bf, Bg)  
**Abb. 1** Screenshot der Bedienoberfläche der Abwasseraufbereitungsanlage mit den Ventilen (V1–V9), den Heizungen (HB1, K1, W1, W2) und den Tanks (Ba, Bb, R1, HB1, Bc, Be, Bd, Bh, Bj, Bk, Bf, Bg)

informed about the purpose of the study and told that they could discontinue participation at any time (in terms of informed consent). All participants were novices in learning the process control task used in the study. The recruitment was similar for all three studies.

Study 2. Eighty students (28 female) took part in the study from April to July 2015. Four participants were excluded based on the selection criteria (Table 2).

Study 3. Eighty students (18 female) took part in the study from October to December 2015. Seven participants were excluded based on the selection criteria (Table 2).

**2.2 Task: Waste water treatment simulation (WaTrSim)**

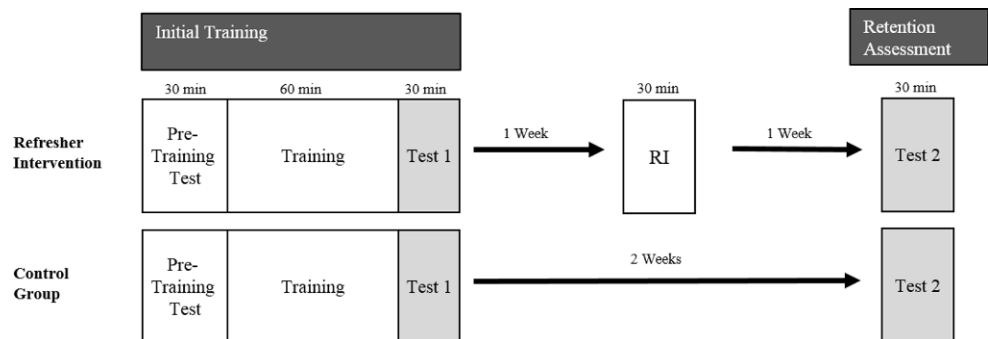
Main task: The complex cognitive skill in the present studies was performed in a simulated process control task embedded in the simulation WaTrSim (Fig. 1; Kluge and Frank 2014).

The participants were trained to execute the particular start-up procedure in a given order and instructed on how to interact with the interface. The procedure comprised the start-up of the plant, which is assumed to be a non-routine task that requires skill retention (Wickens and Hollands 2000). In WaTrSim, the operator’s task is to separate waste water into fresh water and gas by starting up, controlling and monitoring the plant (Kluge and Frank 2014). The operation goal is to maximize the amount of purified gas and to minimize the amount of waste water by executing the start-up procedure in the correct order while considering the right timing for execution. The time permitted to start up the plant is 180s. The start-up procedure differed in all three studies (Fig. 2).

Secondary (monitoring) task: To measure mental workload and to produce a realistic work setting, the participants were required to perform a secondary task in addition to the main task. The secondary task was performed by monitoring the tank level of tank Ba every 50s (Table 2 and Fig. 1;

**Fig. 2** Experimental procedure in study 1–3/RI= Refresher Intervention

**Abb. 2** Experimentelle Prozedur der Studien 1–3/RI= Refresher Intervention





**Table 3** Experimental procedure and time points of the studies**Tab. 3** Experimenteller Verlauf zu den drei Zeitpunkten

Week 1		Week 2	Week 3
Skill acquisition; 120 min		Refresher intervention (for EG only); approx. 30 min	Skill retention; 30 min
Pretraining Test	Socio-demographic data	Practice RI	<b>Test 2</b>
	Retentivity	Skill Test RI	
	Skill acquisition phase		
	2 × explore		5 × start-up (production outcome)
	4 × start-up with manual	Symbolic Rehearsal RI	
Test 1	4 × start-up without manual (production outcome)		

*N* = 266; *RI* refresher intervention, *EG* experimental group

tank Ba can be found on the top left, scoring 0–3 times). The performance measure in this case was the frequency of monitoring the tank level.

Participants were told that their objective was to perform the main and the secondary task in parallel. A video of the two tasks is provided as supplementary material.

**Study 1.** The operation included the start-up procedure of the plant as a fixed-sequence task comprising 13 steps (Fig. 2). Performing the WaTrSim start-up procedure correctly and in a timely manner led to a production outcome of a minimum of 200l of purified gas. The minimum amount of purified gas in the initial training was used as selection criterion ( $\geq 200$ l). The start-up time was max. 180 s.

**Study 2.** The operation included the start-up procedure of the plant as a contingent-sequence task, comprising 13 steps and following four steps for each condition. The following four steps had to be executed depending on the conditions: heating W1  $> 15^\circ\text{C}$  or heating W2  $< 70^\circ\text{C}$ . After one of the conditions had occurred, the correct four steps had to be executed (Fig. 2). Performing the WaTrSim start-up procedure correctly and in a timely manner led to a production outcome of a minimum of 100l of purified gas. The minimum amount of purified gas in initial training was used as selection criterion ( $\geq 100$ l). The start-up time was max. 240 s.

**Study 3.** The operation included the start-up procedure of the plant as a parallel-sequence task. Two sequences had to be operated in parallel: 13 steps for sequence A and three steps for sequence B. Sequence B had to be executed when the level of tank Bf had reached  $> 75\%$  or  $< 25\%$ . After one of the conditions had occurred, the correct two steps had to be executed (Fig. 2). Performing the WaTrSim start-up procedure (both sequences in parallel) correctly and in a timely manner led to a production outcome of a minimum of 200l of purified gas. The minimum amount of purified gas in initial training was used as selection criterion ( $\geq 200$ l). The start-up time was max. 240 s.

## 2.3 Procedure

**Studies 1–3** Participants in the Refresher Conditions operated WaTrSim at three time points (Fig. 2): All participants took part in the initial training (week 1); participants of the experimental group received the refresher intervention one week later (week 2); and after another week, all participants underwent the retention assessment (week 3, Table 3). The control group did not receive a refresher intervention.

**Initial training (Table 3)** The initial training phase lasted for 120 min (Fig. 2). Upon arrival, participants were welcomed and introduced to WaTrSim. After completing tests concerning variables measuring individual differences relevant for the study (general mental ability), participants explored and familiarized themselves with the simulation twice. After this, they trained the start-up procedure (fixed-sequence task, contingent-sequence task or parallel-sequence task) with a manual. Following the training, the participants had to perform the start-up procedure four times without help. They were required to produce a minimum of purified gas (Study 1: 200l, Study 2: 100l, Study 3: 200l). The best trial of this series was used as the reference level of skill mastery after training.

**Refresher intervention** The refresher interventions are described in Table 4. The control group received no refresher intervention.

**Retention assessment (Fig. 3)** Two weeks after the refresher intervention, the retention assessment took place, which lasted for approximately 30 min. After the participants had been welcomed, they were asked to start up the plant. The first trial was used to assess skill retention/decay.

**Studies 1–3** The experimental groups received one of three refresher interventions in week 2 (Practice refresher intervention, Skill Test refresher intervention and Symbolic Re-

Study 1 Fixed-sequence task Start-up procedure: 13 steps		Study 2 Contingent-sequence task Start-up procedure: 13 steps and 2x5 steps		Study 3 Parallel-sequence task Start-up procedure: Sequence A (13 steps) and Sequence B (2x3 steps); both sequences had to be executed in parallel			
Step	Sequence	Step	Sequence	Sequence A		Sequence B	
1	LIC V9: Flow rate 500 l/h	1	LIC V9: Flow rate 500 l/h	1	LIC V9: Flow rate 500 l/h	A	Monitor tank level of tank Bf <i>constantly</i>
2	V2 deactivate follower control	2	V2 deactivate follower control	2	V2 deactivate follower control		↓
3	Valve V1: Flow rate 500 l/h	3	Valve V1: Flow rate 500 l/h	3	Valve V1: Flow rate 500 l/h	B	Tank level Bf > 75%
4	Wait until R1 > 200 l	4	Wait until R1 > 200 l	4	Wait until R1 > 200 l		↓
5	Valve V2: Flow rate 500 l/h	5	Valve V2: Flow rate 500 l/h	5	Valve V2: Flow rate 500 l/h	B	Tank level Bf > 25%
6	Wait until R1 > 400 l	6	Wait until R1 > 400 l	6	Wait until R1 > 400 l	C	FIC V8: Flow rate 90%
7	Valve V3: Flow rate 1000 l/h	7	Valve V3: Flow rate 1000 l/h	7	Valve V3: Flow rate 1000 l/h	C	FIC V8: Flow rate 10%
8	Wait until HB1 > 100 l	8	Wait until HB1 > 100 l	8	Wait until HB1 > 100 l	D	Heating W2: 70°C
9	Activate heating HB1	9	Activate heating HB1	9	Activate heating HB1	D	Heating W2: 20°C
10	Wait until HB1 > 60°C	10	Wait until HB1 > 60°C	10	Wait until HB1 > 60°C		
11	Activate column K1	11	Activate column K1	11	Activate column K1		
12	Valve V4: Flow rate 1000 l/h	12	Valve V4: Flow rate 1000 l/h	12	Valve V4: Flow rate 1000 l/h		
13	Valve V6: Flow rate 400 l/h	13	Valve V6: Flow rate 400 l/h	13	Valve V6: Flow rate 400 l/h		
		14	W1 > 15°C OR W2 > 70°C				
		15	LIC V8 deactivate LIC V8 deactivate				
		16	LIC V9 700 l/h LIC V9 600 l/h				
		17	LIC V8 500 l/h LIC V8 400 l/h				
		18	Heating W1 15°C Heating W2 70°C				

Fig. 3 Description of fixed-sequence task, contingent-sequence task and parallel-sequence task

Abb. 3 Beschreibung der Aufgaben der festgeschriebenen Sequenz, der kontingenten Sequenz und der parallelen Sequenz

hearsal refresher intervention), which are described in Table 4.

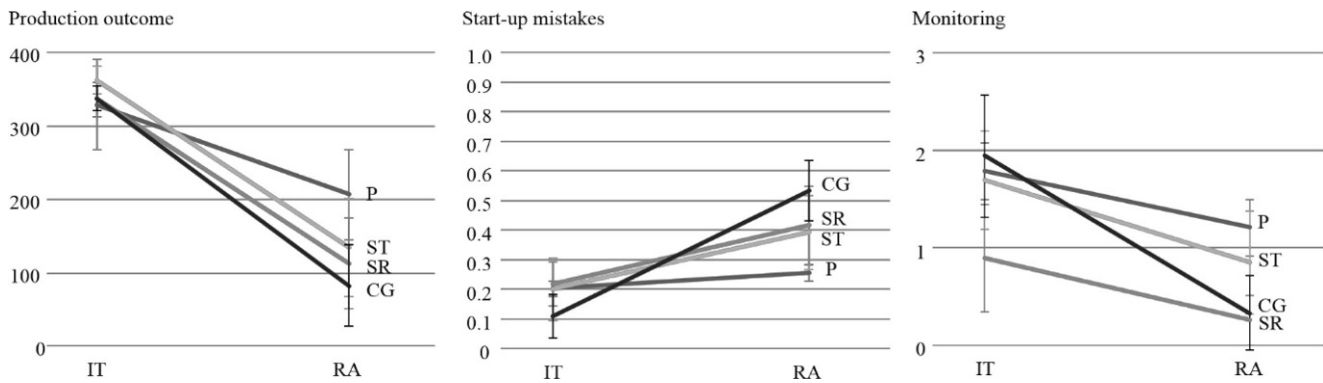
The dependent variables included the effects of the complex cognitive skill execution of the main and secondary task, which are production outcome, start-up mistakes and monitoring. The first trial of retention assessment was used

for the calculations: The production outcome was measured by the produced amount of purified gas at initial training (week 1) and retention assessment (week 3). The minimum production outcome at the initial training varied for each task: Study 1: 200l, Study 2: 100l, Study 3: 200l. The start-up mistakes are the sum of incorrect valve adjustments

Table 4 Description of refresher interventions

Tab. 4 Beschreibung der Refresher Interventionen

Refresher Intervention	Description
Practice	The Practice refresher intervention group executed the start-up procedure of the plant four times and was allowed to use the manual, which included a description of the procedure. The participants were tested in groups of four persons. The intervention took about 30 min.
Skill Test	The Skill Test refresher intervention group was tested individually with only the experimenter in the room and took about 30 min. The participant was given written instructional material in which she/he was asked to imagine that a small town called “Feldkirchen” needs her/his help. The participant was told that she/he is responsible for starting up the plant and producing as much water as possible to save the small town’s water supply. In addition, the participant was explicitly asked to concentrate and focus all her/his attention on the task. It was emphasized that she/he had only one chance to start up the procedure correctly. After this introduction, the participant started up the plant. The instructions stated: “Please concentrate and start up the plant as soon as you feel ready. It is extremely important that the start-up procedure leads to the maximum amount of purified water. The inhabitants of Feldkirchen are depending on you!”
Symbolic Rehearsal	The Symbolic Rehearsal refresher intervention was performed with a computer, took about 30 min to complete, and consisted of the seven symbolic rehearsal tasks: 1) Participants had to fill in the sequence of steps of the start-up procedure, state the flow rate and provide three reasons for producing waste water (instead of purified water). They then had to 2) fill in cloze tasks, 3) arrange steps of the start-up procedure into the correct sequence, and 4) find errors in a presented start-up sequence. 5) Participants had to rehearse the WaTrSim interface, allocate the valve labels in a WaTrSim screenshot and mark the start-up location of the column (K1); 6) they had to rehearse how to operate a valve and a heating by arranging the operating steps into the correct order; and 7) they had had to answer true-or-false questions about how to operate in WaTrSim, e.g. “By clicking the valve, a new dialogue window opens”. After solving the task, participants marked their own results with the help of an answer sheet. All tasks included graphics from WaTrSim. The number of correct answers was used to measure the performance (score: 0–80).



**Fig. 4** Production outcome, start-up mistakes and monitoring performance in study 1. *IT* Initial Training, *RA* Retention Assessment, *P* Practice, *ST* Skill Test, *SR* Symbolic Rehearsal, *CG* Control Group

**Abb. 4** Produktionsleistung, Anfahrfehler und Leistung in der Überwachungsaufgabe in Studie 1

and procedure mistakes, such as adjustment of the incorrect valve flow rate, and also varied for each task: Study 1: 0–15 mistakes, Study 2: 0–19 mistakes, Study 3: 0–21 mistakes (recalculated into percentages, 0–1). Monitoring task level: The secondary task was measured by the number of times the secondary task was performed (Study 1: 0–3, Study 2: 0–4, Study 3: 0–4) by reading out the data from the logfiles.

The control variables were measured according to retentivity and start-up time. Retentivity was measured with the Wilde Intelligence Test-2, which consists of verbal, numerical and figural information (Kersting et al. 2008). First, the participants had to memorize the verbal, numerical and figural information for four minutes. After a disruption phase of 17 min, they then answered reproduction questions related to the memorized information, choosing one of six response options (scores from 0–21; identical for Studies 1–3). The start-up time was measured according to the best production outcome trial of the initial training (week 1) and the first trial of the retention assessment (week 3). The start-up time was limited depending on the task: Study 1: 0–180 s, Study 2: 0–240 s, Study 3: 0–240 s.

### 3 Results

Data from 226 of 240 participants were included in the following calculations. Three participants were excluded in Study 1, four participants were excluded in Study 2 and seven participants were excluded in Study 3. The selection criterion was the produced amount of purified gas as described above. Descriptive statistics are given in Table 5. To ensure that all groups started under the same conditions, the groups were compared regarding the control variables at the initial training. The groups in Studies 1, 2 and 3 were randomized and did not differ significantly in terms of the

control variables ( $p > 0.05$ ), with the exception of retentivity in Study 1 ( $F(3,73) = 5.53$ ,  $p = 0.002$ ,  $\eta_p^2 = 0.19$ ).

### 3.1 Hypothesis-testing

To test the hypotheses, in each study, an ANCOVA with planned contrasts was conducted for the dependent variable production outcome, start-up mistakes and monitoring measured at retention assessment (RA) and covariate production outcome measured at initial training. As Study 1 showed significant differences in retentivity, this was also considered as a covariate.

#### 3.1.1 Study 1—Fixed sequence

**Production outcome:** The analysis regarding production outcome showed a significant difference between groups ( $F(3,71) = 3.27$ ,  $p = 0.026$ ,  $\eta_p^2 = 0.13$ ; Fig. 4). The planned contrasts for analyzing the difference between refresher interventions and control group showed that the refresher interventions supported production outcome significantly higher than no intervention (hypothesis 1:  $p = 0.029$ ). The planned contrasts for analyzing the difference between the Practice and Skill Test refresher intervention showed no difference between the two groups (hypothesis 2:  $p = 0.079$ ). The planned contrast for analyzing the difference between the Skill Test and Symbolic Rehearsal refresher intervention also revealed no difference (hypothesis 3:  $p = 0.900$ ).

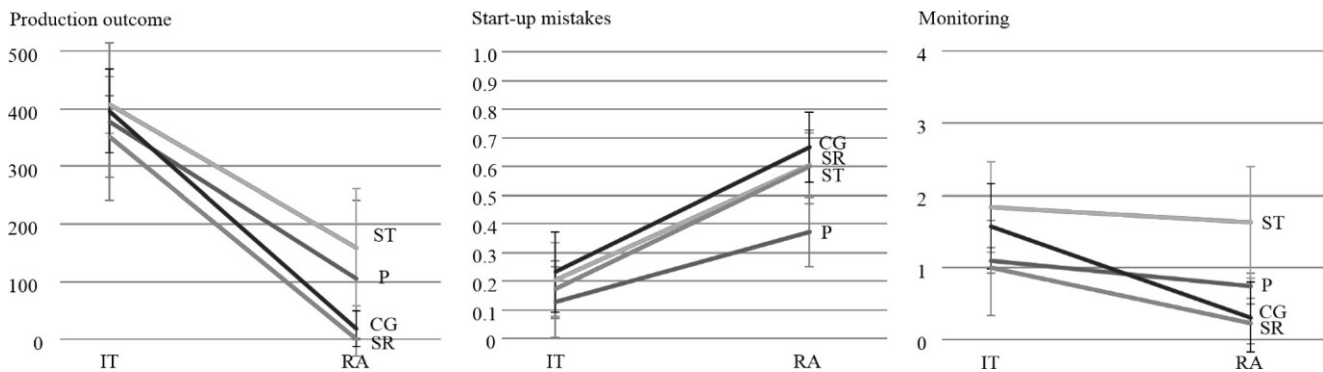
**Start-up mistakes:** The analysis of start-up mistakes showed a significant difference between groups ( $F(3,71) = 5.24$ ,  $p = 0.003$ ,  $\eta_p^2 = 0.18$ ; Fig. 4). The planned contrasts for analyzing the difference between refresher interventions and control group showed that refresher interventions supported the production outcome significantly higher than no intervention (hypothesis 1:  $p = 0.001$ ). The planned contrasts for analyzing the difference between the



**Table 5** Descriptive statistics Means, Standard Deviation and Minimum-Maximum in Brackets  
**Tab. 5** Deskriptive Statistik, Mittelwerte, Standardabweichungen sowie Minimum und Maximum-Werte in Klammern

Control variables	Study 1—fixed-sequence (N=77)				Study 2—contingent-sequence (N=76)				Study 3—parallel-sequence (N=73)			
	Practice	Skill Test	Symbolic Re-hearsal	Control group	Practice	Skill Test	Symbolic Re-hearsal	Control group	Practice	Skill Test	Symbolic Re-hearsal	Control group
Sex	5 female 14 male	7 female 13 male	6 female 14 male	6 female 12 male	10 female 10 male	6 female 13 male	6 female 12 male	6 female 13 male	5 female 15 male	5 female 15 male	6 female 13 male	2 female 12 male
Age	22.47 (3.72, 18–31)	20.90 (2.73, 18–28)	23.55 (3.62, 18–34)	22.33 (2.79, 18–28)	23.50 (2.95, 19–29)	22.11 (3.16, 17–32)	24.06 (3.08, 20–31)	22.63 (3.24, 19–30)	24.35 (4.23, 20–36)	23.80 (2.93, 19–29)	23.10 (3.04, 18–29)	22.20 (3.24, 18–29)
Retentivity	15.37 (1.77, 12–18)	13.00 (3.66, 7–18)	12.65 (2.98, 6–17)	15.33 (1.97, 12–20)	13.80 (3.61, 7–19)	13.79 (3.39, 7–19)	13.50 (3.68, 7–20)	12.74 (3.12, 6–18)	14.40 (2.84, 9–19)	14.75 (2.47, 11–20)	14.20 (3.41, 4–19)	16.36 (2.56, 12–20)
Performance at IT												
IT Start-up time (Study 1: max 180 sec, Studies 2 & 3: max 240 sec)	100.21 (16.21, 77–137)	90.35 (12.87, 71–126)	96.25 (13.57, 73–125)	96.06 (8.48, 80–113)	136.89 (20.27, 95–167)	135.95 (31.11, 74–186)	147.22 (41.35, 79–212)	143.53 (38.77, 92–227)	128.55 (24.22, 93–181)	119.15 (17.10, 93–149)	121.70 (16.11, 100–162)	118.85 (25.12, 93–172)
IT Prod. outcome (Studies 1 & 3: min. 200l, Study 2: min. 100l)	329.39 (50.89, 202.00–411.93)	362.87 (40.08, 275.93–435.97)	150.35 (32.77, 89.00–180.00)	337.46 (33.23, 260.00–399.30)	378.41 (107.49, 132.00–561.41)	407.37 (101.92, 216.00–544.29)	351.15 (143.31, 112.00–526.01)	396.46 (151.38, 112.90–607.89)	435.09 (114.04, 210.11–579.01)	480.65 (78.28, 270.72–610.11)	468.69 (61.36, 311.97–555.01)	489.84 (77.71, 320.36–581.01)
IT Start-up mistakes (0–1)	0.20 (0.23, 0–0.67)	0.20 (0.22, 0–0.73)	0.22 (0.16, 0–0.47)	0.11 (0.15, 0–0.47)	0.13 (0.16, 0–0.53)	0.20 (0.27, 0–0.79)	0.17 (0.2, 0–0.68)	0.25 (0.29, 0–0.84)	0.24 (0.18, 0–0.63)	0.27 (0.22, 0–0.76)	0.20 (0.17, 0–0.53)	0.27 (0.19, 0–0.74)
IT Monitoring (Study 1: 0–3, Studies 2 & 3: 0–4)	1.79 (1.08, 0–3)	1.70 (1.08, 0–3)	0.89 (1.15, 0–3)	1.94 (1.26, 0–3)	1.10 (1.45, 0–4)	1.84 (1.30, 0–4)	1.00 (1.33, 0–4)	1.58 (1.22, 0–4)	1.60 (1.43, 0–4)	1.95 (1.50, 0–4)	1.90 (1.62, 0–4)	3.15 (1.04, 0–4)
Dependent variables at RA												
RA Start-up time	123.32 (35.97, 82–180)	144.40 (36.9, 91–180)	150.35 (32.77, 89–180)	151.72 (31.54, 93–180)	206.80 (45.97, 101–240)	198.53 (50.01, 123–240)	233.28 (28.52, 119–240)	230.89 (30.97, 109–240)	160.10 (45.64, 102–240)	198.20 (46.15, 88–240)	197.00 (40.49, 126–240)	223.95 (28.48, 136–240)
RA Prod. outcome	206.56 (146.59, 0–371.93)	134.6 (144.23, 0–371.93)	113.18 (131.13, 0–363.93)	82.80 (112.08, 0–299.93)	105.82 (156.99, 0–388)	159.72 (211.46, 0–496)	0.63 (2.67, 0–11.33)	18.73 (63.68, 0–273.71)	293.69 (202.56, 0–581.76)	159.25 (159.18, 0–453.80)	157.20 (146.75, 0–431.01)	43.93 (81.61, 0–253.363)
RA Start-up mistakes	0.26 (0.24, 0–0.73)	0.39 (0.27, 0–0.80)	0.42 (0.28, 0–1.00)	0.53 (0.21, 0–0.87)	0.35 (0.29, 0–0.89)	0.60 (0.23, 0.05–0.84)	0.60 (0.26, 0.16–0.95)	0.67 (0.25, 0.16–1.00)	0.35 (0.22, 0–0.74)	0.48 (0.31, 0–0.86)	0.43 (0.20, 0.05–0.81)	0.64 (0.23, 0.06–0.86)
RA Monitoring	1.21 (1.18, 0–3)	0.85 (1.14, 0–3)	0.25 (0.55, 0–2)	0.33 (0.77, 0–3)	0.75 (1.33, 0–4)	1.63 (1.61, 0–4)	0.22 (0.55, 0–2)	0.32 (1.00, 0–4)	1.45 (1.47, 0–4)	0.80 (1.32, 0–4)	0.35 (0.75, 0–3)	0.90 (1.29, 0–4)

IT initial training, RA retention assessment



**Fig. 5** Production outcome, start-up mistakes and monitoring performance in study 2. *IT* Initial Training, *RA* Retention Assessment, *P* Practice, *ST* Skill Test, *SR* Symbolic Rehearsal, *CG* Control Group

**Abb. 5** Produktionsleistung, Anfahrfehler und Leistung in der Überwachungsaufgabe in Studie 2

Practice and Skill Test refresher intervention showed no difference between the two groups ( $p=0.138$ ).

The planned contrast for analyzing the difference between the Skill Test and Symbolic Rehearsal refresher intervention showed no difference (hypothesis 3:  $p=0.872$ ).

Monitoring: The analysis of monitoring showed a significant difference between groups ( $F(3,70)=4.56$ ,  $p=0.006$ ,  $\eta^2_p=0.16$ ; Fig. 4). The planned contrasts for analyzing the difference between refresher interventions and control group showed that refresher interventions supported the production outcome significantly higher than no intervention (hypothesis 1:  $p=0.004$ ). The planned contrasts for analyzing the difference between the Practice and Skill Test refresher intervention showed no difference between the two groups (hypothesis 2:  $p=0.238$ ). The planned contrast for analyzing the difference between the Skill Test and Symbolic Rehearsal refresher intervention also showed no difference (hypothesis 3:  $p=0.472$ ).

In conclusion, the results of Study 1 indicate that a refresher intervention supports performance better than no intervention. However, Practice and Skill Test refresher interventions support performance equally well. There was no difference in performance between the Skill Test refresher intervention and the Symbolic Rehearsal refresher intervention.

### 3.1.2 Study 2—Contingent sequence.

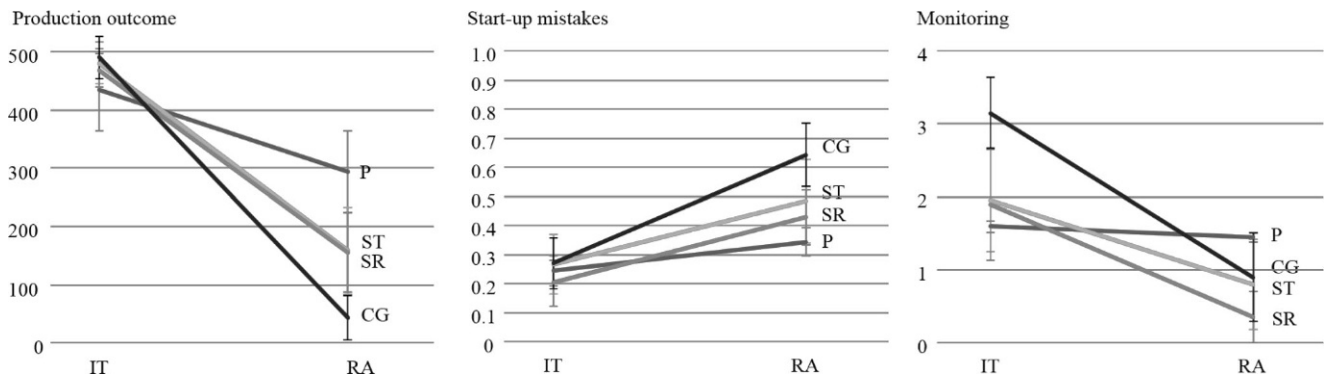
Production outcome: The analysis regarding production outcome showed a significant difference between groups ( $F(3,67)=5.09$ ,  $p=0.003$ ,  $\eta^2_p=0.18$ ; Fig. 5). The planned contrasts for analyzing the difference between refresher interventions and control group showed no superior performance of refresher interventions (hypothesis 1:  $p=0.074$ ). The planned contrasts for analyzing the difference between the Practice and Skill Test refresher intervention showed no difference between the two groups (hypothesis 2:  $p=0.247$ ). The planned contrast for analyzing the difference between

the Skill Test and Symbolic Rehearsal refresher intervention showed a significantly superior performance of the Skill Test refresher intervention (hypothesis 3:  $p=0.001$ ).

Start-up mistakes: The analysis of start-up mistakes showed a significant difference between groups ( $F(3,67)=4.25$ ,  $p=0.008$ ,  $\eta^2_p=0.16$ ; Fig. 5). The planned contrasts for analyzing the difference between refresher interventions and control group showed no superior performance of refresher interventions (hypothesis 1:  $p=0.085$ ). The planned contrasts for analyzing the difference between the Practice and Skill Test refresher intervention showed a significantly superior performance of the Practice refresher intervention (hypothesis 2:  $p=0.007$ ). The planned contrast for analyzing the difference between the Skill Test and Symbolic Rehearsal refresher intervention showed no difference (hypothesis 3:  $p=0.900$ ).

Monitoring: The analysis of monitoring showed a significant difference between groups ( $F(3,70)=4.80$ ,  $p=0.004$ ,  $\eta^2_p=0.17$ ; Fig. 5). The planned contrasts for analyzing the difference between refresher interventions and control group showed that a refresher intervention supports production outcome significantly higher than no intervention (hypothesis 1:  $p=0.028$ ). The planned contrasts for analyzing the difference between the Practice and Skill Test refresher intervention showed no difference between the two groups (hypothesis 2:  $p=0.100$ ). The planned contrast for analyzing the difference between the Skill Test and Symbolic Rehearsal refresher intervention showed a significantly superior performance of the Skill Test refresher intervention (hypothesis 3:  $p=0.004$ ).

The results of Study 2 demonstrate that refresher interventions support the performance of the secondary task (monitoring) better than no intervention. However, the refresher interventions did not show a superior performance with regard to production outcome and start-up mistakes. Additionally, the Practice and Skill Test refresher interventions supported performance equally well in terms of production outcome and monitoring, but Practice supported the



**Fig. 6** Production outcome, start-up mistakes and monitoring performance in study 3. *IT* Initial Training, *RA* Retention Assessment, *P* Practice, *ST* Skill Test, *SR* Symbolic Rehearsal, *CG* Control Group

**Abb. 6** Produktionsleistung, Anfahrfehler und Leistung in der Überwachungsaufgabe in Studie 3

performance of start-up mistakes better. A comparison of the Symbolic Rehearsal refresher intervention and the Skill Test refresher intervention revealed a superior effect of the latter for production outcome and monitoring.

### 3.1.3 Study 3—Parallel sequence.

**Production outcome:** The analysis regarding production outcome showed a significant interaction of time and group ( $F(3,68)=12.10$ ,  $p<0.001$ ,  $\eta_p^2=0.35$ ; Fig. 6). The planned contrasts for analyzing the difference between refresher interventions and control group showed that a refresher intervention supports production outcome significantly higher than no intervention (hypothesis 1:  $p<0.001$ ). The planned contrasts for analyzing the difference between the Practice and Skill Test refresher intervention showed a significantly superior performance of the Practice refresher intervention (hypothesis 2:  $p<0.001$ ). The planned contrast for analyzing the difference between the Skill Test and Symbolic Rehearsal refresher intervention showed no difference (hypothesis 3:  $p=0.858$ ).

**Start-up mistakes:** The analysis regarding production outcome showed a significant interaction of time and group ( $F(3,68)=4.80$ ,  $p=0.004$ ,  $\eta_p^2=0.18$ ; Fig. 6). The planned contrasts for analyzing the difference between refresher interventions and control group showed that a refresher intervention supports production outcome significantly higher than no intervention (hypothesis 1:  $p=0.001$ ). The planned contrasts for analyzing the difference between the Practice and Skill Test refresher intervention showed no difference between the two groups (hypothesis 2:  $p=0.082$ ). The planned contrast for analyzing the difference between the Skill Test and Symbolic Rehearsal refresher intervention showed no difference (hypothesis 3:  $p=0.716$ ).

**Monitoring:** The analysis regarding production outcome showed a significant interaction of time and group ( $F(3,68)=4.21$ ,  $p=0.009$ ,  $\eta_p^2=0.16$ ; Fig. 6). The planned

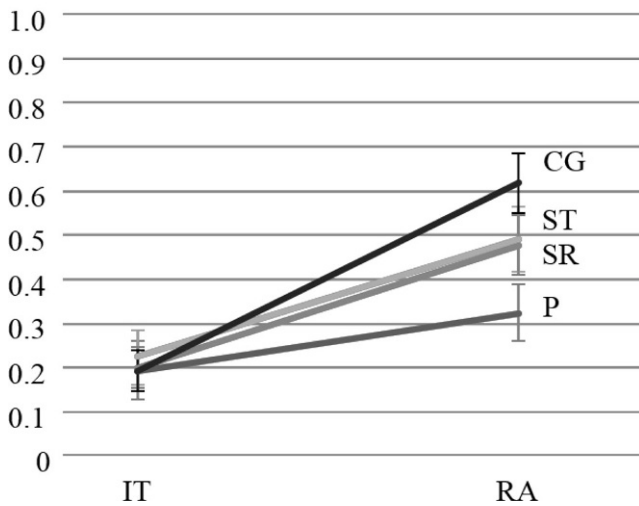
contrasts for analyzing the difference between refresher interventions and control group showed no difference between the two groups (hypothesis 1:  $p=0.469$ ). The planned contrasts for analyzing the difference between the Practice and Skill Test refresher intervention showed a significantly superior performance of the Practice refresher intervention (hypothesis 2:  $p=0.028$ ). The planned contrast for analyzing the difference between the Skill Test and Symbolic Rehearsal refresher intervention showed no difference (hypothesis 3:  $p=0.235$ ).

In conclusion, the results of Study 3 suggest that refresher interventions support performance better with regard to production outcome and start-up mistakes. Additionally, the Practice refresher intervention supported performance better than the Skill Test refresher intervention with respect to production outcome and monitoring; the two groups performed equally well regarding start-up mistakes. The comparison of the Skill Test refresher intervention and the Symbolic Rehearsal refresher intervention revealed a superior effect of the former with regard to production outcome and monitoring.

### 3.1.4 Comparison refresher interventions independent of task

Finally, the data from all three studies were combined and the groups were compared. As production outcome requirements, start-up time and monitoring differed between the three studies, the only measure which permitted a comparison between all studies was the percentage of start-up mistakes. The analysis regarding production outcome showed a significant interaction of time and group ( $F(3,218)=12.54$ ,  $p<0.001$ ,  $\eta_p^2=0.15$ ; Fig. 7). The planned contrasts for analyzing the difference between refresher interventions and control group showed that a refresher intervention supports production outcome significantly higher than no intervention (hypothesis 1:  $p<0.001$ ). The

Start-up mistakes



**Fig. 7** Comparison of Refresher-Intervention independent of task regarding start up mistakes. *IT* Initial Training, *RA* Retention Assessment, *P* Practice, *ST* Skill Test, *SR* Symbolic Rehearsal, *CG* Control Group

**Abb. 7** Vergleich von Refresher-Interventionen unabhängig vom Aufgabentyp

planned contrasts for analyzing the difference between the Practice and Skill Test refresher intervention showed a significantly superior performance of the Practice refresher intervention (hypothesis 2:  $p=0.001$ ). The planned contrast for analyzing the difference between the Skill Test and Symbolic Rehearsal refresher intervention showed no difference (hypothesis 3:  $p=0.790$ ).

This indicates that each refresher intervention supports performance with regard to start-up mistakes better than no intervention. Moreover, a Practice refresher intervention supports an error-free performance better than a Skill Test refresher intervention. However, no superior effect of Practice and Skill Test was found compared to Symbolic Rehearsal.

The results of the three studies and the overall analysis are summarized in Table 6, ordered according to the hypotheses.

**4 Discussion**

The aim of the present paper was to find the best way to support skill retention. We therefore compared the effect of a Practice, Skill Test and Symbolic Rehearsal refresher intervention in a fixed-, a contingent-, and a parallel-sequence task. All three tasks consisted of a main task (production outcome, start-up mistakes) and a secondary task (monitoring). To summarize, we found the following: Hypothesis 1: Refresher interventions were more effective than no intervention in fixed- and parallel-sequence tasks with regard to the main task, but refresher interventions were not superior to no intervention in the contingent-sequence task. With regard to the secondary task, refresher interventions supported performance in the fixed- and the contingent-sequence task. The overall comparison suggests that refresher interventions are better able to support performance of the main task compared to no intervention. Hypothesis 2: When analyzing the suggested superior effect of a Practice refresher intervention over a Skill Test refresher intervention, a superior performance of the Practice refresher intervention was only found for the performance of start-up mistakes in the contingent-sequence task and for the performance of production outcome and monitoring in the parallel-sequence task. The performance in the fixed-sequence task was supported equally well by the Practice and the Skill Test refresher intervention, as was the performance regarding production outcome and monitoring in the contingent-sequence task, and the performance regarding start-up mistakes in the parallel-sequence task.

The overall comparison indicates that the Practice refresher intervention supported an error-free performance

**Table 6** Summary of study results for the hypotheses  
**Tab. 6** Zusammenfassung der Ergebnisse bezogen auf die Hypothesen

	Study 1	Study 2	Study 3	Studies 1–3
Hypothesis 1: Refresher intervention > control group				
Main task: Production outcome	✓ $p=0.029$	✗	✓ $p<0.001$	–
Main task: Start-up mistakes	✓ $p=0.001$	✗	✓ $p=0.001$	✓ $p<0.001$
Secondary task: Monitoring	✓ $p=0.004$	✓	$p=0.028$ ✗	–
Hypothesis 2: Practice > Skill Test				
Main task: Production outcome	✗	✗	✓ $p<0.001$	–
Main task: Start-up mistakes	✗	✓	$p=0.007$ ✗	✓ $p=0.001$
Secondary task: Monitoring	✓	✗	✓ $p=0.028$	–
Hypothesis 3: Skill Test > Symbolic Rehearsal				
Main task: Production outcome	✗	✓	$p=0.001$ ✗	–
Main task: Start-up mistakes	✗	✗	✗	✗
Secondary task: Monitoring	✗	✓	$p=0.004$ ✗	–

better than the Skill Test refresher intervention. In discussions with colleagues from the field of instructional psychology (Frank et al. 2016), we also heard arguments that the testing effect seems more likely to occur with static learning material, and less likely to occur when refreshing dynamic tasks, as is required in the operation of WaTrSim. Hypothesis 3: Moreover, the comparison of the Skill Test and the Symbolic Rehearsal refresher intervention indicated that the former did not support the performance better than the latter in the fixed-sequence task. Nevertheless, in the contingent-sequence task, the Skill Test refresher intervention supported the performance better in both the main and secondary task. However, the overall comparison showed no superior effect of the Skill Test over the Symbolic Rehearsal refresher intervention or the Practice refresher as is assumed by scholars promoting the testing effect. That means the refresher interventions in general were shown to be important, but the impact and effectiveness of a specific refresher instruction varies throughout the concrete task, and its cognitive demands. E. g. the three sequences differed with respect to length and attention management (e. g. attention sharing in the parallel task). The more the task required the remembrance of steps (in the contingent sequence), the more effective was Skill testing. The more important the aspect of attention management (in the parallel sequence) the more effective was practice, as only further practice enables the operator to practice the attention sharing demands. In conclusion further research should address the multiple cognitive and information processing task demands and their unique emphasis in relation to successful task execution. A cognitive task analysis with an emphasis on memory requirements of sub tasks is assumed to be a promising basis for designing refresher intervention. In that respect, combinations of refresher interventions and their distribution over time can compensate for the shortcomings of a single instruction applied alone.

#### 4.1 Limitations

Although it may be criticized that the results of three different studies were analyzed to compare the effect of the developed refresher interventions, it should be noted that the study procedures and designs are comparable and differ only with respect to the task types. Nevertheless, it would be valuable to conduct a study with extended retention intervals in order to create a more realistic setting.

So far, the results are limited to a prototypical cognitive task which is found in process control and is skill- and ruled-based. Although sequences are also found in many other occupations, e. g. aviation, shipping and many other commercial or military fields, further evidence and empirical work is needed to find the one best way to support skill retention, if indeed there is one. Moreover, for the time be-

ing, the generalizability of our results is limited to periods of weeks of non-use, and we cannot transfer our findings to tasks which comprise months or years of retention intervals. Finally, our participants were all novices. It is assumed that ongoing or continuing job experience attenuates skill loss at least a little. However, job experience does not fully attenuate skill loss (Casner et al. 2013; Kluge et al. 2014), as it is known that time on task is more important than tenure.

#### 4.2 Implications

Implications for further research, but also for practical applications, can be derived from the limitations: Future research should include different tasks, longer periods of non-use and experienced operators. Further challenges are the possible interaction between concrete on-the-job experiences in daily operations and skill retention, the interaction between person-related variables, on-the-job experiences and skill retention, as well as reasonable combinations and sequences of refresher interventions. New measurement and intervention methods, such as experience sampling methods implemented on mobile devices, could be used to track the actually performed on-the-job tasks of operators over longer periods of time and relate them to target performance measures which require skill retention.

As skill retention might be a lifelong challenge, evidence-based rules for combining and alternating refresher interventions should be investigated. These might also be implemented, for example, with the help of computer-based methods or mobile devices.

In this respect, our three studies made a first attempt to compare refresher intervention effects systematically, but the many different work settings which were not addressed in the present work leave plenty of scope for more research to support lifelong skill retention.

### 5 Conclusion

To summarize, there is no one best way to support skill retention, but it appears that a skill test refresher intervention might be the most worthwhile: Although it is not superior to practice, it does seem to be more efficient in some cases (in the more difficult contingent-sequence task), as it requires less time to apply. Nevertheless, if companies are keen to use skill tests, they should keep in mind that they might come at a cost, for instance more start-up mistakes and possibly a higher workload.

In conclusion, refresher interventions and the issue of skill retention should receive as much attention as initial training and learning processes in vocational and occupational settings. We do not know as much about skill retention as we know about training for instance summarized



by Bell et al. (2017). Especially for organizations that expect lifelong remembering in combination with only infrequently required skills, for instance due to a high level of automation in production, demands of worker flexibility and a high task variety, or long periods of non-use. Efforts for further research can support evidence-based decisions on the design of refresher interventions that fit the organizations' and task needs' in terms of effectiveness and efficiency best.

**Funding** This research was funded by German Research Foundation (KL2207/3–3).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Anderson JR (1982) Acquisition of cognitive skill. *Psychol Rev* 89(4):369–406
- Annett J (1979) Memory for skill. In: Gruneberg MM, Morris PE (eds) *Applied problems in memory*. Academic Press, London, pp 213–247
- Arthur W, Day EA (2013a) Introduction: knowledge and skill decay in applied research. In: Arthur W, Day EA, Bennett W, Portrey AM (eds) *Individual and team skill decay*. Routledge, New York, pp 3–13
- Arthur W, Day EA (2013b) The past, present, and future of applied skill decay research. In: Arthur W, Day EA, Bennett W, Portrey AM (eds) *Individual and team skill decay*. Routledge, New York, pp 3–13
- Arthur W, Bennett W, Stanush PL, McNelly TL (1998) Factors that influence skill decay and retention: a quantitative review and analysis. *Hum Perf* 11(1):57–101
- Arthur W, Day EA, Villado AJ, Boatman PR, Kowolik V, Bennett W, Bhupatkar A (2010) The effect of distributed practice on immediate posttraining, and long-term performance on a complex command-and-control simulation task. *Hum Perf* 23(5):428–445
- Bainbridge L (1983) Ironies of automation. *Automatica (Oxf)* 19(6):775–779
- Bell BS, Tannenbaum SI, Ford JK, Noe RA, Kraiger K (2017) 100 years of training and development research: what we know and where we should go. *J Appl Psychol* 102(3):305–323
- Bjork RA (2009) Structuring the conditions of training to achieve elite performance: reflections on elite training programs and related themes in chapters 10–13. In: Ericsson KA (ed) *Development of professional expertise. Toward measurement of expert performance and design of optimal learning environments*. Cambridge University Press, Cambridge, pp 312–332
- Bjork RA (2011) On the symbiosis of remembering, forgetting, and learning. In: Benjamin AS (ed) *Successful remembering and successful forgetting: a Festschrift in honor of Robert A. Bjork*. Psychology Press, London, pp 1–22
- Bjork RA, Bjork EL (1992) A new theory of disuse and an old theory of stimulus fluctuation. In: Healy A, Kosslyn S, Shiffrin R (eds) *Essays in honor of William K. Estes, vol. 1: from learning theory to connectionist theory; vol. 2: from learning processes to cognitive processes*. Lawrence Erlbaum, Hillsdale, pp 35–67
- Bjork RA, Bjork EL (2006) Optimizing treatment and instruction: implications of a new theory of disuse. In: Nilsson LG, Ohta N (eds) *Memory and society. Psychological perspectives*. Psychology Press, New York, pp 116–140
- Bodilly SJ, Fernandez J, Kimbrough J, Purnell S (1986) Individual ready reserve skill retention and refresher training options. Rand Corporation, Santa Monica
- Brehmer B (1992) Dynamic decision making: human control of complex systems. *Acta Psychol (Amst)* 81(3):211–241
- Casner SM, Geven RW, Williams KT (2013) The effectiveness of airline pilot training for abnormal events. *Human Factors* 55(3):477–485
- Childs JM, Spears WD, Prophet WW (1983) Private pilot flight skill retention 8, 16, and 24 month following certification. Federal Aviation Administration, Washington
- Cooper G, Tindall-Ford S, Chandler P, Sweller J (2001) Learning by imagining. *J Exp Psychol Appl* 7(1):68–82
- Dörner D, Bick T (1994) *Lohhausen: Vom Umgang mit Unbestimmtheit und Komplexität*. Huber, Bern
- Driskell JE, Copper C, Moran A (1994) Does mental practice enhance performance? *J Appl Psychol* 79(4):481–492
- Fanjoy RO, Keller JC (2013) Flight skill proficiency issues in instrument approach accidents. *J Aviat Technol Eng* 3(1):17–23
- Farmer E, van Rooij J, Riemersma J, Jorna P, Moraal J (1999) *Handbook of simulator-based training*. Ashgate, Aldershot
- Farr MJ (1987) *The long-term retention of knowledge and skills*. Routledge, New York
- Faul F, Erdfelder E, Lang AG, Buchner A (2007) G\*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39:175–191
- Foss MA, Fabiani M, Mané AM, Donchin E (1989) Unsupervised practice: the performance of the control group. *Acta Psychol (Amst)* 71(1–3):23–51
- Frank B, Lau M, Kluge A (2016) The effect of testing complex cognitive skill retention in a fixed sequence task, a dual task and decision making task. 9th International Cognitive Load Theory Conference, Bochum, June 22nd–24th 2016
- Ginzburg S, Dar-El EM (2000) Skill retention and relearning—a proposed cyclical model. *J Workplace Learn* 12(8):327–332
- van Gog T, Sweller J (2015) Not new, but nearly forgotten: the testing effect decreases or even disappears as the complexity of learning materials increases. *Educ Psychol Rev* 27(2):247–264
- van Gog T, Kester L, Dirx K, Hoogerheide V, Boerboom J, Verkoijen P (2015) Testing after worked example study does not enhance delayed problem-solving performance compared to restudy. *Educ Psychol Rev* 27(2):265–289
- Gonzalez C, Lerch JF, Lebiere C (2003) Instance-based learning in dynamic decision making. *Cogn Sci* 27(4):591–635
- Hager W, Hasselhorn M (2008) Pädagogisch-psychologische Interventionsmaßnahmen. In: Schneider W, Hasselhorn M (eds) *Handbuch der Pädagogischen Psychologie*. Hogrefe, Göttingen, pp 339–347
- Hagman JD, Rose AM (1983) Retention of military tasks: a review. *Hum Factors* 25(2):199–213
- Hoffman R, Feltovich P, Fiore S, Klein G, Missildine W, DiBello L (2010) Accelerated proficiency and facilitated retention: recommendations based on an integration of research and findings from a working meeting. Florida Institute for Human and Machine Computing, Pensacola
- Jaber MY, Kher HV (2004) Variant versus invariant time to total forgetting: the learn-forget curve model revisited. *Comput Ind Eng* 46(4):697–705
- Karpicke JD, Aue WR (2015) The testing effect is alive and well with complex material. *Educ Psychol Rev* 27(2):317–326
- Karuppan CM (2011) Learning and forgetting: implications for workforce flexibility in AMT environments. In: Jaber MY (ed) *Learning curves: theory, models, and applications*. CRC, Boca Raton, pp 173–190

- Kaye W, Mancini ME, Rallis SF (1987) Advanced cardiac life support refresher course using standardized objective-based mage code testing. *Crit Card Med* 15(1):55–60
- Kersting M, Althoff K, Jäger AO (2008) Wilde-Intelligenz-Test 2 (WIT-2). Hogrefe, Göttingen
- Kim JW, Ritter FE, Koubek RJ (2013) An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theor Issues Ergon Sci* 14(1):22–37
- King BF (2015) An analysis of federal aviation administration knowledge test scores and fatal general aviation accidents. <http://krex.k-state.edu/dspace/handle/2097/19024>. Accessed 17 March 2016
- Kluge A (2014) The acquisition of knowledge and skills for taskwork and teamwork to control complex technical systems. Springer, Heidelberg
- Kluge A, Frank B (2014) Counteracting skill decay: four refresher interventions and their effect on skill and knowledge retention in a simulated process control task. *Ergonomics* 57(2):175–190
- Kluge A, Burkolter D, Frank B (2012) “Being prepared for the infrequent”: a comparative study of two refresher training approaches and their effects on temporal and adaptive transfer in a process control task. *Proc Hum Factors Ergon Soc Annu Meet* 56:2437–2441
- Kluge A, Frank B, Maafi S, Kuzmanovska A (2015) Does skill retention benefit from reactivity and symbolic rehearsal?—two studies with a simulated process control task. *Ergonomics* 13(1):1–16
- Kluge A, Frank B, Miebach J (2014) Measuring skill decay in process control—results from four experiments with a simulated process control task. In: Waard D, Brookhuis K, Wiczorek R, Di Nocera F, Barham P, Weikert C, Kluge A, Gerbino W, Toffetti A (eds) *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference*, pp 79–93 (<http://conference.hfes-europe.org/>)
- Kluge A, Nazir S, Manca D (2014) Advanced applications in process control and training needs of field and control room operators. *IIE Trans Occup* 2(3–4):121–136
- Kluge A, Sauer J, Schüler K, Burkolter D (2009) Designing training for process control simulators: a review of empirical findings and current practices. *Theor Issues Ergon Sci* 10(6):489–509
- Kontogiannis T, Shepherd A (1999) Training conditions and strategic aspects of skill transfer in a simulated process control task. *Hum Comput Interact* 14(4):355–393
- Leahy W, Hanham J, Sweller J (2015) High element interactivity information during problem solving may lead to failure to obtain the testing effect. *Educ Psychol Rev* 27(2):291–304
- Mattoon JS (1994) Designing instructional simulations: effects of instructional control and type of training task on developing display-interpretation skills. *Int J Aviat Psychol* 4(3):189–209
- McDaniel MA, Roediger HL, McDermott KB (2007) Generalizing test-enhanced learning from the laboratory to the classroom. *Psychon Bull Rev* 14(2):200–206
- Moray N (1996) A taxonomy and theory of mental models. *Proc Hum Factors Ergon Soc Annu Meet* 40(4):164–168
- Morris NM, Rouse WB (1985) Review and evaluation of empirical research in troubleshooting. *Hum Factors* 27:503–530
- Naylor JC, Briggs GE (1961) Long-term retention of learned skills: a review of the literature. *ASD Tech Rep* 61(390):35
- Nembhard DA, Uzumeri MV (2000) Experiential learning and forgetting for manual and cognitive tasks. *Int J Ind Ergon* 25(4):315–326
- Omerod TC, Shephard A (2004) Using task analysis for information requirements specification: the sub-goal template method. In: Diaper D, Stanton NA (eds) *The handbook of task analysis for human-computer interaction*. CRC, Boca Raton, pp 347–365
- Pashler H, Rohrer D, Cepeda NJ, Carpenter SK (2007) Enhancing learning and retarding forgetting: choices and consequences. *Psychon Bull Rev* 14(2):187–193
- Perez RS, Skinner A, Weyhrauch P, Niehaus J, Lathan C, Schweitzberg SD, Cao CG (2013) Prevention of surgical skill decay. *Mil Med* 178(10S):76–86
- Proctor RW, Dutta A (1995) *Skill acquisition and human performance*. SAGE, Thousand Oaks
- Roediger HL, Karpicke JD (2006) The power of testing memory: basic research and implications for educational practice. *Perspect Psychol Sci* 1(3):181–210
- Sauer J, Burkolter D, Kluge A, Ritzmann S, Schüler K (2008) The effects of heuristic rule training on operator performance in a simulated process control environment. *Ergonomics* 51(7):953–967
- Schendel JD, Hagman JD (1982) On sustaining procedural skills over a prolonged retention interval. *J Appl Psychol* 67(5):605–610
- Schneider W (1985) Training high-performance skills: fallacies and guidelines. *Hum Factors* 27(3):285–300
- Schumacher EH, Seymour TL, Glass JM, Fencsik DE, Lauber EJ, Kieras DE, Meyer DE (2001) Virtually perfect time sharing in dual-task performance: uncorking the central cognitive bottleneck. *Psychol Sci* 12(2):101–108
- Stammers RB (1981) Skill retention and control room tasks. *Proc Hum Factors Ergon Soc Annu Meet* 25(1):168–172
- Sun R, Merrill E, Peterson T (2001) From implicit skills to explicit knowledge: a bottom-up model of skill learning. *Cogn Sci* 25:203–244
- Sun R, Slusarz P, Terry C (2005) The interaction of the explicit and the implicit in skill learning: a dual-process approach. *Psychol Rev* 112(1):159
- Villado AJ, Day EA, Arthur W, Boatman PR, Kowolik V, Bhupatkar A, Bennett W (2013) Use of, reaction to, and efficacy of observation rehearsal training: enhancing skill retention on a complex command-and-control simulation task. In: Arthur W, Day EA, Bennett W, Portrey AM (eds) *Individual and team skill decay*. Routledge, New York, pp 240–257
- Walley S (1995) Mental rehearsal techniques for reducing skill decay of unemployed commercial pilots. In: Johnston N, Fuller R, McDonald N (eds) *Aviation psychology: training and selection*. Ashgate, Aldershot, pp 215–221
- Wang X, Day EA, Kowolik V, Schuelke MJ, Hughes MG (2013) Factors influencing knowledge and skill decay in organizational training: a meta-analysis. In: Arthur W, Day EA, Bennett W, Portrey AM (eds) *Individual and team skill decay*. Routledge, New York, pp 68–117
- Wickens CD (2008) Multiple resources and mental workload. *Hum Factors* 50(3):449–455
- Wickens CD, Hollands JG (2000) *Engineering psychology and human performance*. Prentice Hall, Upper Saddle River
- Wickens CD, McCarley JS (2008) *Applied attention theory*. CRC, Boca Raton
- Wollard M, Whitfield R, Newcombe RG, Colquhoun M, Vetter N, Chamberlain D (2006) Optimal refresher training intervals for AED and CPR skills: a randomised controlled trial. *Resuscitation* 71(2):237–247