

AFRL-AFOSR-UK-TR-2022-0021

Sparse sampling-based estimation of the retention of rarely used procedural knowledge

van Rijn, D.H.
RIJKSUNIVERSITEIT GRONINGEN
BROERSTRAAT 5
GRONINGEN, GRONINGEN, 9712 CP
NLD

02/04/2022
Final Technical Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
Air Force Office of Scientific Research
European Office of Aerospace Research and Development
Unit 4515 Box 14, APO AE 09421

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
20220204	Final	START DATE 20151115	END DATE 20191114

4. TITLE AND SUBTITLE
Sparse sampling-based estimation of the retention of rarely used procedural knowledge

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
	FA9550-16-1-0043	
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
D.H. van Rijn

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
RIJKSUNIVERSITEIT GRONINGEN
BROERSTRAAT 5
GRONINGEN, GRONINGEN 9712 CP
NLD

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
EOARD UNIT 4515 APO AE 09421-4515	AFRL/AFOSR IOE	AFRL-AFOSR-UK-TR-2022-0021

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
In this project, we assessed whether cognitive measures can be used to predict performance on Cardiopulmonary Resuscitation (CPR), a complex, life-saving task. More specifically, we
collected detailed performance measures during the (re-)learning of CPR and on a number of carefully selected cognitive tasks, and correlated these with CPR performance months to years
after the initial training. We assessed CPR performance in a group of German young adults who all had received CPR training (as part of their driver's licence exam) months to years prior to
their participation in the study. Even though their initial CPR performance was far below the required proficiency levels, a quick, video-based retraining increased CPR performance to well above minimal proficiency levels. Based on these findings, we have published guidelines and a recommendation for regular, low-cost video-based refresher training sessions to keep CPR performance of the general population at reasonable levels (Maass et al., 2019). During this initial session, we also collected performance data on a number of traditional, cognitive laboratory tasks that were selected based on their relation to aspects of the CPR task. For example, we assessed participants' capacity to follow a rhythm similar to the compression frequency during CPR, or their capacity to memorize simple facts as CPR training also involved the memorization of certain procedures and the parameters of these procedures. Our goal was to assess whether such measures predict (1) CPR performance, and (2) the decline in CPR performance over time. As this resulted in an uniquely rich dataset of both laboratory tasks and a well-described, real-life task, we have ensured that this data will be available to the scientific community at-large by depositing all data and an extensive description of this data in an open access scientific "data-paper" (Sense et al., 2019). We have also collected CPR and additional cognitive task data 6 months after the initial data collection, and 1.5 years after initial data collection, and have made all these datasets available for further analysis. More specifically, we have been in close contact with the Air Force Research Laboratory to ensure that the data collected by us could be used in future computational modeling projects led by the AFRL on predicting skilled performance over extended time ranges.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	SAR	71

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
NANDINI IYER	314-235-6161

Final Report of the Project:

Sparse sampling-based estimation of the retention of rarely used procedural knowledge

PI: Hedderik van Rijn, Florian Sense, Sarah Maaß
University of Groningen
The Netherlands

Award no. FA9550-16-1-0043, Air Force Office of Scientific Research

Project/Grant Period:

November 15, 2015 to November 14, 2019

Project Focus:

Although the last details are still the focus of extensive scientific discussion, at a more global level the acquisition of declarative, factual knowledge and its subsequent fading over time are well described. However, how procedural knowledge is retained, and then specifically procedural knowledge that largely depends on a combination of procedural skills and script knowledge describing what to do when, is less well studied. This is partly due to the difficulty finding well controlled, non-artificial tasks that can be used in the laboratory. Moreover, in many procedural tasks, retention intervals are often measured in weeks or months, rather than in minutes or hours as in declarative knowledge. In this project, we propose to validate CPR-training as a task that taps into procedural learning and to demonstrate that it can be used in laboratory settings to assess the parameters that determine successful script-based skill-learning. (Text taken from the initial Technical Proposal)

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Project Summary:

In this project, we assessed whether cognitive measures can be used to predict performance on Cardiopulmonary Resuscitation (CPR), a complex, life-saving task. More specifically, we collected detailed performance measures during the (re-)learning of CPR and on a number of carefully selected cognitive tasks, and correlated these with CPR performance months to years after the initial training. We assessed CPR performance in a group of German young adults who all had received CPR training (as part of their driver's licence exam) months to years prior to their participation in the study. Even though their initial CPR performance was far below the required proficiency levels, a quick, video-based retraining increased CPR performance to well above minimal proficiency levels. Based on these findings, we have published guidelines and a recommendation for regular, low-cost video-based refresher training sessions to keep CPR performance of the general population at reasonable levels (Maass et al., 2019).

During this initial session, we also collected performance data on a number of traditional, cognitive laboratory tasks that were selected based on their relation to aspects of the CPR task. For example, we assessed participants' capacity to follow a rhythm similar to the compression frequency during CPR, or their capacity to memorize simple facts as CPR training also involved the memorization of certain procedures and the parameters of these procedures. Our goal was to assess whether such measures predict (1) CPR performance, and (2) the decline in CPR performance over time. As this resulted in an uniquely rich dataset of both laboratory tasks and a well-described, real-life task, we have ensured that this data will be available to the scientific community at-large by depositing all data and an extensive description of this data in an open-access scientific "data-paper" (Sense et al., 2019).

We have also collected CPR and additional cognitive task data 6 months after the initial data collection, and 1.5 years after initial data collection, and have made all these datasets available for further analysis. More specifically, we have been in close contact with the Air Force Research Laboratory to ensure that the data collected by us could be used in future computational modeling projects led by the AFRL on predicting skilled performance over extended time ranges.

Project Outcomes:

The main project outcomes are documented in the two journal publications in the four conference presentations given during the project (see Project Output). Moreover, the year reports provide more detailed information about the activities per year.

- CPR performance, even after months or years of initial training can easily be restored to certification levels by watching a short refresher video (Maass et al, 2019).
- Predicting performance on a complex task such as CPR is complex, and we have not been able to identify simple, cognitive laboratory tasks that predict CPR performance

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

with practically, meaningful accuracy. Even though the absence of evidence does not imply evidence of absence, and thus we cannot rule out that other cognitive tasks would predict fluctuations in CPR performance, there are a number of considerations that render a simple match between cognitive lab-tasks and CPR performance as evaluated by current standards unlikely:
    - 1. CPR performance is measured as an interplay between declarative and procedural knowledge, whereas standard laboratory tasks typically assess either declarative, or procedural knowledge. It might well be necessary to explicitly test for the interaction between these two strands of knowledge.
    - 2. CPR performance depends partially on a number of binary variables, for example, whether or not assistance is requested. It is difficult to find reliable relationships given the binary nature of these variables.
    - 3. Performing CPR requires a different mindset than participating in computer-based, “innocent” cognitive tasks. Even though the impact of performing CPR on a manikin will be less than needing to perform actual CPR, it might well be that personality traits will play a role in CPR performance, something which is not captured in the laboratory tasks.
- CPR performance in the general population months to years after initial training or retraining is highly variable, ranging from still (well) above minimum performance levels, to even failing to engage in any CPR activity. This further strengthens the argument that personalized training schemes would provide a large improvement in safety *and* efficiency as some individuals would require even more regular refresher training sessions than the current default, whereas for others a refresher training at longer intervals might suffice.

Recommendations for Future Projects:

- Adaptive training that is optimized to individual learning characteristics could provide significant learning gains.
- CPR consists of procedural and declarative skills. Training and testing these skills separately could provide further optimization, as the declarative skills could easily be tested in an online testing environment.

Collaborations / Meetings:

Van Rijn, Sense, and Maaß visited the Gluck’s research group at the Air Force Research Laboratory in Dayton. OH from November 9th to 11th, 2016; and from July 17th to 19th, 2018. During the first meeting, we designed and set up the main studies based on input derived from earlier work done by the AFRL on complex skill acquisition (see Oermann, et al., 2020). During the second meeting, we planned the final round of data collection and discussed how the data

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

collected by the Groningen team could be used for refining computational models of learning and performance that are developed by the AFRL. (In addition, the members of the AFRL and Groningen teams met at international conferences, such as the ICCM in Coventry, UK). Based on the interactions during these meetings, Sense was offered a continuing (external) research position at AFRL, and Van Rijn and Sense have started a continuing collaboration with Jastrzembski and colleagues of the AFRL on optimization of memorization (e.g., Sense et al., 2019).

Oermann, M. H., Krusmark, M. A., Kardong-Edgren, S., Jastrzembski, T. S., & Gluck, K. A. (2020). Training interval in cardiopulmonary resuscitation. *Plos one*, *15*(1), e0226786.

Sense, F., Jastrzembski, T., Mozer, M.C., Krusmark, M., & van Rijn, H. (2019). Perspectives on Computational Models of Learning and Forgetting. Proceedings of the 17th International Conference on Cognitive Modeling. Montreal, Canada.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Output:

Journal Publications:

Sense, F., Maaß, S., Gluck, K., & van Rijn, H. (2019). Within-subject performance on a real-life, complex task and traditional lab experiments: Measures of word learning, Raven matrices, tapping, and CPR. *Journal of cognition*, *2*(1).

- https://www.journalofcognition.org/articles/10.5334/joc.65/

Maaß, S. C., Sense, F., Gluck, K., & Van Rijn, H. (2019). Keeping Bystanders Active: Resuscitating Resuscitation Skills. *Frontiers in Public Health*, *7*, 177.

- https://www.frontiersin.org/articles/10.3389/fpubh.2019.00177/full

International Conference Contributions:

Sense, F., Maaß, S.C., & Van Rijn, H. (2016). Interactions of declarative and procedural memory in real-life tasks: validating CPR as a new paradigm. In D. Reitter & F. E. Ritter (Eds.), Proceedings of the 14th International Conference on Cognitive Modeling (pp. 252–253). University Park, PA: Penn State.

- http://acs.ist.psu.edu/iccm2016/proceedings/sense2016iccm.pdf

Maaß, S.C., Sense, F., & Van Rijn, H. (2017). Cognitive Modeling of Cardiopulmonary Resuscitation Knowledge and Skill Spanning Months to Years (abstract). In M. K. van Vugt, A. P. Banks, & W.G. Kennedy (Eds.), Proceedings of the 15th International Conference on Cognitive Modeling (pp. 228–229). Coventry, UK: University of Warwick.

- http://mathpsych.org/conferences/2017/file/MP_ICCM2017_Abstract_Booklet.pdf

Maaß, S., Sense, F., Krusmark, M., Gluck, K. A., & van Rijn, H. (2019). Deconvolving a Complex, Real-Life Task: Do standard lab tasks predict CPR learning and retention?. In *CogSci* (p. 3520).

- https://cogsci.mindmodeling.org/2019/papers/0828/index.html

Gluck, K.A., Collins, M., Krusmark, M., Sense, F., Maass, S.C., & van Rijn, H. (2019). Predicting Performance in Cardiopulmonary Resuscitation. *Proceedings of the 17th International Conference on Cognitive Modeling*. Montreal, Canada.

- https://iccm-conference.neocities.org/2019/proceedings/papers/ICCM2019_paper_38.pdf

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Below follow the three yearly reports and the associated appendixes, each preceded by an empty page.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Year One Project Report
Project: "Sparse sampling-based estimation of the retention of rarely used procedural knowledge."

Principle Investigator: Prof. dr. H. van Rijn
Department of Experimental Psychology and Department of Psychometrics and Statistics
University of Groningen, The Netherlands

Grant number: FA9550-16-1-0043

Period of performance: November 2015- November 2016

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Abstract

Although in many real-life settings it is of utmost importance to quickly learn and retain skills that depends on a combination of procedural skills and script knowledge describing what to do when, this type of learning is rarely studied. For example, professionals need to learn Cardiopulmonary resuscitation (CPR) and refresh their knowledge periodically to remain certified. Work performed by dr. Gluck (AFRL) has demonstrated that individual differences can be observed in how quickly individuals learn and forget CPR. However, current procedure is that everyone is re-trained at the same intervals. This is not optimal: some people do not train often enough and forget the skill potentially leading to detrimental outcomes, whereas others train too often, wasting resources. The aim of this project is to develop a system that accurately predicts when an individual should be re-trained to maintain optimal retention of a complex skill, more specifically of CPR.
To be able to conduct experiments with valid and reliable measurements the laboratory was equipped with a Laerdal Resusci Anne QCPR manikin (see Figure 1) and a Laerdal SimPad SkillReporter.



Figure 1: Laerdal QCPR Manikin

In following experiments participants were to be taught “adult basic life support” skills in accordance with the guidelines of the European Resuscitation Council (Perkin et al., 2015). Hereto, participants were to be trained by watching an instructional video. This video was made specifically in the scope of this project and is made available on the Open Science Foundation. To test the equipment and set-up of a future experiment a pilot study was run in April 2016. The aim was to gather insight on which aspects of CPR performance relate to individual variations as estimated in laboratory tasks of procedural learning and fact learning. The latter relation provides information of whether the system we have developed to improve fact learning based on an adaptive learning algorithm (Sense et al., 2016) can be used for skill learning and to predict future skilled performance. The results of this pilot study were presented as a poster on the International Conference on Cognitive Modeling in Penn State (Sense, Maass, & Van Rijn, 2016; see attachments for poster and conference paper). In November 2016 a visit to the laboratory of dr. Gluck (AFRL), Ohio, was scheduled for a collaborative project meeting to share past experience and results and design subsequent experiments for 2017. The outcome of this meeting was a detailed experimental procedure for the first experiment conducted in this project.
In the following parts of this overview we describe the video (Attachment 1), the poster presented at the ICCM 2016 (Attachment 2), the conference paper discussing the pilot experiment (Attachment 3) and the experimental procedure for Experiment 1 to be conducted in 2017 (Attachment 4).

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Attachment 1:

The instructional video is made available on the Open Science Framework (https://mfr.osf.io/render?url=https://osf.io/9mw24/?action=download%26mode=render).





Figure 2. Screenshots taken from the instructional video

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Attachment 2: Poster ICCM 2016

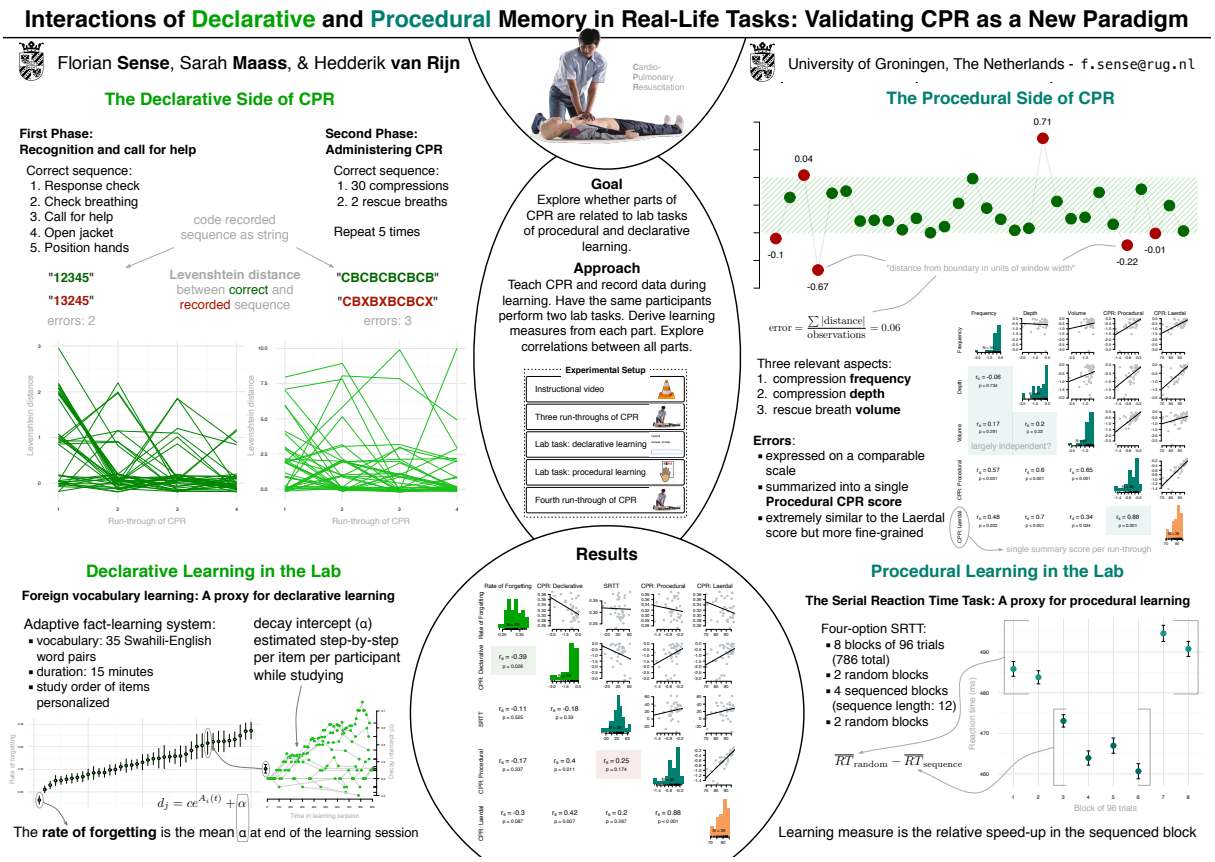


Figure 3. Poster presented at the ICCM 2016

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Attachment 3: Conference Paper ICCM 2016

Interactions of Declarative and Procedural Memory in Real-Life Tasks: Validating CPR as a New Paradigm

Florian Sense (f.sense@rug.nl)
Department of Experimental Psychology and Department of Psychometrics and Statistics,
Groningen, The Netherlands

Sarah Maass (s.c.maass@rug.nl)
Research School of Behavioral and Cognitive Neuroscience,
Groningen, The Netherlands

Hedderik van Rijn (d.h.van.rijn@rug.nl)
Department of Experimental Psychology and Department of Psychometrics and Statistics,
Groningen, The Netherlands

Keywords: declarative memory; procedural memory; complex skill; new paradigm

Introduction

In the learning and memory literature, there is a clear distinction between procedural and declarative memory. We know that they develop differently in children (Finn et al., 2016) and that they are dissociated anatomically and this distinction is reflected in cognitive architectures as well (Anderson et al., 2004). In the lab, most tasks tap into and measure either one of those components. In most complex real-life skills, however, both declarative and procedural memory are required to perform well.

Here we will look at the learning of a complex real-life skill: cardiopulmonary resuscitation (CPR). More specifically, we will teach participants *basic life support* skills meant to be performed on an adult victim suffering from cardiac arrest. CPR has both declarative and procedural components and there are clear guidelines prescribing the sequence of steps that need to be executed (Perkins et al., 2015). This sequence needs to be remembered (e.g., *I need to call am ambulance* before *I initiate CPR*) which draws on declarative memory. Learning to administer correct compressions, on the other hand, is probably closer to a procedural skill.

CPR is an ideal task for various reasons. It is clearly constrained and can be performed and monitored in a controlled (lab) environment while staying close to how it would be trained in a real-life setting. Publicly accessible guidelines provide clear learning criteria against which the obtained measures can be compared. The skill can be learned in a single session and pilot data suggest that there are individual differences in how quickly CPR performance decreases after initial learning and that not all aspects of CPR are retained equally well. Therefore, CPR is an ideal testbed to investigate the interaction between declarative and procedural learning in a real-life setting. In this exploratory study, we will investigate which aspects of CPR performance are best predicted by someone's declarative learning or procedural learning ability.

Tasks

To address the research question, we will have each participant perform three tasks. Each task is intended to measure one of the three components of interest: acquisition and performance of CPR, procedural learning, and declarative learning.

Learning CPR

Each participant is taught "adult basic life support" skills in accordance with the guidelines of the European Resuscitation Council (Perkins et al., 2015). Hereto, participants watch an instructional video and then practice CPR with a Laerdal Resusci Anne QCPR manikin. Data is recorded using the Laerdal SimPad SkillReporter. This setup allows detailed recordings of the skill development during the acquisition of basic life support skills: both the order of the steps that were taken can be recorded as well as detailed measures of each compression and rescue breath that is administered. After an initial training phase (with corrective feedback), CPR performance will be assessed about 45 minutes later and either one or four weeks later.

Procedural Learning

To assess a participant's expertise in procedural learning, we selected separate implicit and explicit procedural knowledge tasks.

Serial Reaction Time Task. In the serial reaction time task (SRTT) visual cues appear over four response options and the participant needs to respond with one of four fingers that are mapped to the response options. The task can be used as a measure of implicit procedural learning by presenting the participant with different blocks of trials: one block in which the order of responses is random and one in which the order is a repeating sequence (Robertson, 2007). The participant is (usually) not aware of the sequence but their reaction times become markedly faster. A learning measure can be computed by subtracting the mean reaction time in the sequenced block from the mean in the random block (e.g., Willingham, Salidis, & Gabrieli, 2002).

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

sees a line with 12 corners connecting two points and needs to trace the line. However, they can only see the line and their own hand through a mirror. With practice, both the completion time and the number of errors decreases, indicating procedural learning. Quantifying the improvement provides a learning measure that is linked to individual differences (e.g., Finn et al., 2016).

Declarative Learning

To derive a learning measure for someone's ability to learn declarative information, we use a fact learning system developed in our lab. The system uses retrieval practice to quiz learners on a trial-by-trial basis. This allows recordings of accuracy and response latencies which are used to estimate the current memory strength of the test item, which are used to schedule a repetition of the item before it is forgotten (Van Rijn, van Maanen, & van Woudenberg, 2009). As more information is gathered, an estimate of how quickly an item is forgotten is fine-tuned. After studying 35 Swahili-English word-pairs for 15 minutes, an estimated rate of forgetting can be obtained for each participant. This learning measure indicates how quickly, on average, a participant forgets this type of material. We have shown that this measure can be measured reliably (Sense, Behrens, Meijer, & Van Rijn, 2016) and is not related to common measures of executive attentional functioning (Sense, Meijer, & Van Rijn, accepted).

Procedure

A total of 40 participants with no prior CPR experience will be invited to participate. So far, data from 20 participants has been collected. Participants will be invited for two sessions. In the first, they will be taught CPR and will complete the 15-minute word-learning session, 768 trials of the SRTT (half random, half sequenced), and nine mirror-tracing trials. At the end of session one, they will be tested on the CPR performance. Either one or four weeks later, in session two, participants will be tested on the Swahili words they learned as well as on their CPR performance. Also, they will complete another round of 768 SRTT and nine mirror-tracing trials.

Planned Analyses

This is an exploratory study to verify whether the acquisition and forgetting of CPR skills can be captured in this paradigm and how CPR performance is related to established declarative and procedural learning tasks.

The computation of learning measures for the declarative and procedural tasks is straightforward, however, valid measures for CPR still need to be developed. As a large number of measures are recorded, the main challenge will be to construct suitable learning measures. We will report a first exploration of various options and discuss their merits.

The derived learning measures can then be correlated with the learning measures for the declarative and procedural tasks. Multiple regression can be used to determine which (combination of) measures can best explain variance in overall CPR performance. Furthermore, we will look at which aspects of CPR are forgotten more quickly than others and lead to a decrease in overall CPR performance.

Hopefully, this project will provide useful information with regards to CPR's relationship to declarative and procedural learning as well as indications that might be helpful to optimize (re-)learning of CPR. Furthermore, steps will be taken to establish CPR as a complex real-life task that involves both declarative and procedural components. As such, modeling human behavior in this task can shed more light on the relative contributions of declarative knowledge and procedural skills in complex human behavior.

Acknowledgements

This work was supported by EOARD grant 11926121. EOARD was not involved in planning or reporting this study.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–60.

Finn, A. S., Kalra, P. B., Goetz, C., Leonard, J. A., Sheridan, M. A., & Gabrieli, J. D. E. (2016). Developmental dissociation between the maturation of procedural memory and declarative memory. *Journal of Experimental Child Psychology*, *142*, 212–220.

Perkins, G. D., Handley, A. J., Koster, R. W., Castrén, M., Smyth, M. A., Olasveengen, T., … Greif, R. (2015). European Resuscitation Council Guidelines for Resuscitation 2015. Section 2. Adult basic life support and automated external defibrillation. *Resuscitation*, *95*, 81–99.

Robertson, E. M. (2007). The Serial Reaction Time Task: Implicit Motor Skill Learning? *The Journal of Neuroscience*, *27*(38), 10073–10075.

Sense, F., Behrens, F., Meijer, R. R., & Van Rijn, H. (2016). An Individual's Rate of Forgetting is Stable over Time, but Differs Across Materials. *Topics in Cognitive Science*, *8*(1), 305–321.

Sense, F., Meijer, R. R., & Van Rijn, H. (accepted). On the Link between Fact Learning and General Cognitive Ability. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*.

Van Rijn, H., van Maanen, L., & van Woudenberg, M. (2009). Passing the test: Improving learning gains by balancing spacing and testing effects. In *Proceedings of the 9th International Conference on Cognitive Modeling* (pp. 110–115).

Willingham, D. B., Salidis, J., & Gabrieli, J. D. E. (2002). Direct comparison of neural systems mediating conscious and unconscious skill learning. *Journal of Neurophysiology*, *88*(3), 1451–1460.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Attachment 4: Experimental Procedure for Experiment 1

The acquisition and retention of factual and procedural knowledge is impacted by multiple factors. Most basically, as material is studied, memory performance improves. The change in performance approximates a power law – additional practice produces further improvement, but at a diminishing rate (Newell & Rosenbloom, 1981). Conversely, memory performance degrades during periods of non-use. This also approximates a power law, with rapid loss occurring initially followed by more-gradual, sustained loss (Anderson & Schooler, 1991; Rubin & Wenzel, 1996). In addition to these factors, knowledge acquisition and retention is impacted by the temporal distribution of practice. Separating practice repetitions by a delay slows learning but enhances retention (Cepeda et al., 2006).

The Predictive Performance Equation (PPE; Jastrzembski & Gluck, 2009; Walsh, Gluck, Gunzelmann, Jastrzembski, & Krusmark, in revision) is an extension of the General Performance Equation (GPE; Anderson & Schunn, 2000). Both PPE and GPE represent how amount of practice and elapsed time since test impact learning and retention. However, PPE also accounts for how the temporal distribution of practice affects retention. In PPE, including more space between practice repetitions reduces decay rate, thereby increasing retention. In the research reported here, we test the ability of PPE to generate subject-level performance predictions for cardiopulmonary resuscitation (CPR). Such predictions, if valid, could be used to prescribe personalized refresher training intervals.

In addition to testing a model of knowledge acquisition and retention, we sought to study learning and memory in conditions more reflective of the real world. Although there is an extensive scientific literature associated with learning, forgetting, and spacing effects, much of it suffers from two deficits in ecological validity. First, most published studies of these effects involve abstract, simplified laboratory tasks. Second, in most of these studies, all of the learning and retention taking place in short time periods, often within a single experiment session.

Participants are to be German citizens enrolled as students at the University of Groningen. A requirement for initial issuance of a driver's license in Germany is the successful completion of first aid training that includes CPR. Thus, each participant in this study completed initial CPR training when they received their driver's license. The initial study session shall start with one bout of CPR to assess current performance levels, after which participants are retrained and complete three further CPR training bouts (with a 10 min delay between the initial test and the next two bouts, and then an hour between the 3th and 4th sessions). In addition to CPR data, we will also collect Raven-based IQ scores, fact learning proficiency using the SlimStampen algorithm (Sense, Behrens, Meijer, & van Rijn, 2016), and data from a Serial Reaction Time Task (Robertson, 2007).

During the experiment session, CPR performance scores ranging from 0 to 100 are automatically generated and recorded by the manikin for each training bout. We will calibrate PPE separately to each of 50 participants based on their training history and performance scores. We do not know how participants scored during the initial training they completed before receiving their driver's license but we know the approximate delay since the training. To include this observation in a participant's training history

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

which is the threshold for adequate performance. We then used the calibrated model to predict retention performance after an 18-week delay.

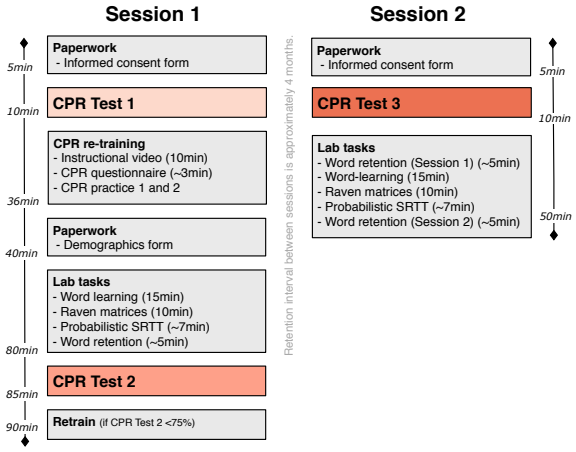


Figure 4: Experimental set-up

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

References

Anderson, J. R., & Schunn, C. (2000). Implications of the ACT-R learning theory: No magic bullets. *Advances in instructional psychology, Educational design and cognitive science*, 1-33.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological science*, *2*(6), 396-408.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380. http://doi.org/10.1037/0033-2909.132.3.354

Jastrzembski, T. S., & Gluck, K. A. (2009). A formal comparison of model variants for performance prediction. *Proceedings of the International Conference on Cognitive Modeling*, Manchester, UK.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive skills and their acquisition*, *1*, 1-55.

Perkins, G. D., Handley, A. J., Koster, R. W., Castrén, M., Smyth, M. A., Olasveengen, T., … Greif, R. (2015). European Resuscitation Council Guidelines for Resuscitation 2015. Section 2. Adult basic life support and automated external defibrillation. Resuscitation , 95 , 81–99.

Robertson, E. M. (2007). The serial reaction time task: implicit motor skill learning? *Journal of Neuroscience*, *27*(38), 10073-10075.

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103*, 734–760.

Sense, F., Behrens, F., Meijer, R. R. and van Rijn, H. (2016). An Individual's Rate of Forgetting Is Stable Over Time but Differs Across Materials. Topics in Cognitive Science, 8: 305–321.

Sense, F., Maass, S., & Rijn, H. van. (2016). Interactions of declarative and procedural memory in real-life tasks: validating CPR as a new paradigm (abstract). In D. Reitter & F. E. Ritter (Eds.), Proceedings of the 14th International Conference on Cognitive Modeling (pp. 252–253). University Park, PA: Penn State.

Walsh, M. W., Gluck, K. A., Gunzelmann, G., Jastrzembski, T., & Krusmark, M. (in revision). Evaluating the theoretical adequacy and applied potential of computational models of the spacing effect.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Sparse Sampling-Based Estimation of the Retention of Rarely Used Procedural Knowledge

van Rijn, Sense & Maass

Although in many real-life settings it is of utmost importance to quickly learn (and retain) skills that depends on a combination of procedural skills and script knowledge describing what to do when, this type of learning is rarely studied.

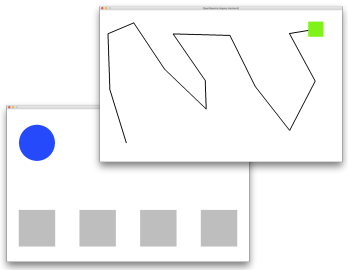
- For example, professionals need to learn CPR and refresh their knowledge periodically to remain certified.
- Work performed by dr. Gluck (AFRL) has demonstrated that individual differences can be observed in how quickly individuals learn and forget CPR.
- However, current procedure is that everyone is re-trained at the same intervals
- This is not optimal: some people do not train often enough and forget the skill, whereas others train too often, wasting resources.

university of groningen

Can we develop a system that accurately predicts when an individual should be re-trained to maintain optimal retention of a complex skill?

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

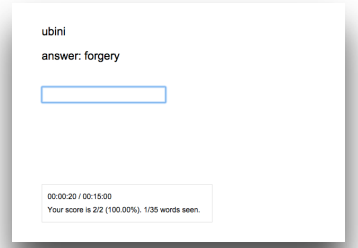
Current Status: Data Collection for 1st Experiment Started



predict?



predict?

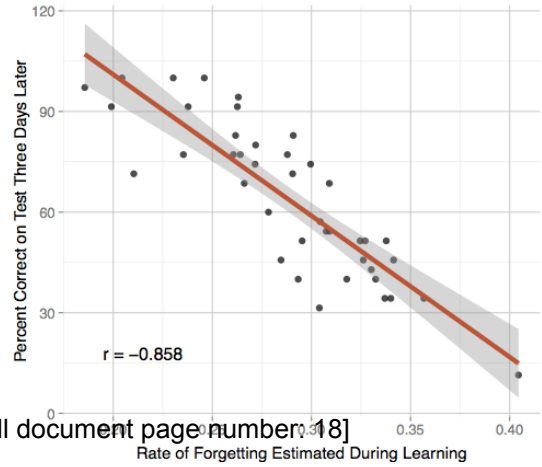


Measures of Procedural Learning

Learning CPR

Measures of Declarative/Fact Learning

- CPR is a complex skill that involves many components.
- The first experiment will provide information on which aspects of CPR performance relate to individual variation as estimated in:
  - laboratory tasks of procedural learning
  - laboratory tasks of fact learning
- Subsequent experiments will be designed in collaboration with dr. Gluck (AFRL).
- These studies will provide information on whether the system we have developed to improve fact learning based on an adaptive learning algorithm can be used (1) for skill learning, and (2) to *predict future skilled performance*.



[Full document page number: 18]

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Predictions & Outcomes

- The work in this project will tell us more about
  - the link between declarative and procedural learning and how these memory systems interact to form complex skills.
  - how the learning and retention of complex skills develops over time.
  - how we can track the learning and forgetting of complex skills over time to predict when performance will drop below a predetermined level.

- Being able to predict the time point at which a skill will be forgotten has important practical implications. Training of complex skills, such as CPR, for example, could be made more effective.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Interactions of Declarative and Procedural Memory in Real-Life Tasks: Validating CPR as a New Paradigm

Florian Sense (f.sense@rug.nl)
Department of Experimental Psychology and Department of Psychometrics and Statistics,
Groningen, The Netherlands

Sarah Maass (s.c.maass@rug.nl)
Research School of Behavioral and Cognitive Neuroscience,
Groningen, The Netherlands

Hedderik van Rijn (d.h.van.rijn@rug.nl)
Department of Experimental Psychology and Department of Psychometrics and Statistics,
Groningen, The Netherlands

Keywords: declarative memory; procedural memory; complex skill; new paradigm

Introduction

In the learning and memory literature, there is a clear distinction between procedural and declarative memory. We know that they develop differently in children (Finn et al., 2016) and that they are dissociated anatomically and this distinction is reflected in cognitive architectures as well (Anderson et al., 2004). In the lab, most tasks tap into and measure either one of those components. In most complex real-life skills, however, both declarative and procedural memory are required to perform well.

Here we will look at the learning of a complex real-life skill: cardiopulmonary resuscitation (CPR). More specifically, we will teach participants *basic life support* skills meant to be performed on an adult victim suffering from cardiac arrest. CPR has both declarative and procedural components and there are clear guidelines prescribing the sequence of steps that need to be executed (Perkins et al., 2015). This sequence needs to be remembered (e.g., *I need to call am ambulance* before *I initiate CPR*) which draws on declarative memory. Learning to administer correct compressions, on the other hand, is probably closer to a procedural skill.

CPR is an ideal task for various reasons. It is clearly constrained and can be performed and monitored in a controlled (lab) environment while staying close to how it would be trained in a real-life setting. Publicly accessible guidelines provide clear learning criteria against which the obtained measures can be compared. The skill can be learned in a single session and pilot data suggest that there are individual differences in how quickly CPR performance decreases after initial learning and that not all aspects of CPR are retained equally well. Therefore, CPR is an ideal testbed to investigate the interaction between declarative and procedural learning in a real-life setting. In this exploratory study, we will investigate which aspects of CPR performance are best predicted by someone's declarative learning or procedural learning ability.

Tasks

To address the research question, we will have each participant perform three tasks. Each task is intended to measure one of the three components of interest: acquisition and performance of CPR, procedural learning, and declarative learning.

Learning CPR

Each participant is taught "adult basic life support" skills in accordance with the guidelines of the European Resuscitation Council (Perkins et al., 2015). Hereto, participants watch an instructional video and then practice CPR with a Laerdal Resusci Anne QCPR manikin. Data is recorded using the Laerdal SimPad SkillReporter. This setup allows detailed recordings of the skill development during the acquisition of basic life support skills: both the order of the steps that were taken can be recorded as well as detailed measures of each compression and rescue breath that is administered. After an initial training phase (with corrective feedback), CPR performance will be assessed about 45 minutes later and either one or four weeks later.

Procedural Learning

To assess a participant's expertise in procedural learning, we selected separate implicit and explicit procedural knowledge tasks.

Serial Reaction Time Task. In the serial reaction time task (SRTT) visual cues appear over four response options and the participant needs to respond with one of four fingers that are mapped to the response options. The task can be used as a measure of implicit procedural learning by presenting the participant with different blocks of trials: one block in which the order of responses is random and one in which the order is a repeating sequence (Robertson, 2007). The participant is (usually) not aware of the sequence but their reaction times become markedly faster. A learning measure can be computed by subtracting the mean reaction time in the sequenced block from the mean in the random block (e.g., Willingham, Salidis, & Gabrieli, 2002).

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Mirror Tracing. In the mirror tracing task, the participant sees a line with 12 corners connecting two points and needs to trace the line. However, they can only see the line and their own hand through a mirror. With practice, both the completion time and the number of errors decreases, indicating procedural learning. Quantifying the improvement provides a learning measure that is linked to individual differences (e.g., Finn et al., 2016).

Declarative Learning

To derive a learning measure for someone's ability to learn declarative information, we use a fact learning system developed in our lab. The system uses retrieval practice to quiz learners on a trial-by-trial basis. This allows recordings of accuracy and response latencies which are used to estimate the current memory strength of the test item, which are used to schedule a repetition of the item before it is forgotten (Van Rijn, van Maanen, & van Woudenberg, 2009). As more information is gathered, an estimate of how quickly an item is forgotten is fine-tuned. After studying 35 Swahili-English word-pairs for 15 minutes, an estimated rate of forgetting can be obtained for each participant. This learning measure indicates how quickly, on average, a participant forgets this type of material. We have shown that this measure can be measured reliably (Sense, Behrens, Meijer, & Van Rijn, 2016) and is not related to common measures of executive attentional functioning (Sense, Meijer, & Van Rijn, accepted).

Procedure

A total of 40 participants with no prior CPR experience will be invited to participate. So far, data from 20 participants has been collected. Participants will be invited for two sessions. In the first, they will be taught CPR and will complete the 15-minute word-learning session, 768 trials of the SRTT (half random, half sequenced), and nine mirror-tracing trials. At the end of session one, they will be tested on the CPR performance. Either one or four weeks later, in session two, participants will be tested on the Swahili words they learned as well as on their CPR performance. Also, they will complete another round of 768 SRTT and nine mirror-tracing trials.

Planned Analyses

This is an exploratory study to verify whether the acquisition and forgetting of CPR skills can be captured in this paradigm and how CPR performance is related to established declarative and procedural learning tasks.

The computation of learning measures for the declarative and procedural tasks is straightforward, however, valid measures for CPR still need to be developed. As a large number of measures are recorded, the main challenge will be to construct suitable learning measures. We will report a first exploration of various options and discuss their merits.

The derived learning measures can then be correlated with the learning measures for the declarative and procedural tasks. Multiple regression can be used to determine which (combination of) measures can best explain variance in overall CPR performance. Furthermore, we will look at which aspects of CPR are forgotten more quickly than others and lead to a decrease in overall CPR performance.

Hopefully, this project will provide useful information with regards to CPR's relationship to declarative and procedural learning as well as indications that might be helpful to optimize (re-)learning of CPR. Furthermore, steps will be taken to establish CPR as a complex real-life task that involves both declarative and procedural components. As such, modeling human behavior in this task can shed more light on the relative contributions of declarative knowledge and procedural skills in complex human behavior.

Acknowledgements

This work was supported by EOARD grant 11926121. EOARD was not involved in planning or reporting this study.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–60.

Finn, A. S., Kalra, P. B., Goetz, C., Leonard, J. A., Sheridan, M. A., & Gabrieli, J. D. E. (2016). Developmental dissociation between the maturation of procedural memory and declarative memory. *Journal of Experimental Child Psychology*, *142*, 212–220.

Perkins, G. D., Handley, A. J., Koster, R. W., Castrén, M., Smyth, M. A., Olasveengen, T., … Greif, R. (2015). European Resuscitation Council Guidelines for Resuscitation 2015. Section 2. Adult basic life support and automated external defibrillation. *Resuscitation*, *95*, 81–99.

Robertson, E. M. (2007). The Serial Reaction Time Task: Implicit Motor Skill Learning? *The Journal of Neuroscience*, *27*(38), 10073–10075.

Sense, F., Behrens, F., Meijer, R. R., & Van Rijn, H. (2016). An Individual's Rate of Forgetting is Stable over Time, but Differs Across Materials. *Topics in Cognitive Science*, *8*(1), 305–321.

Sense, F., Meijer, R. R., & Van Rijn, H. (accepted). On the Link between Fact Learning and General Cognitive Ability. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*.

Van Rijn, H., van Maanen, L., & van Woudenberg, M. (2009). Passing the test: Improving learning gains by balancing spacing and testing effects. In *Proceedings of the 9th International Conference on Cognitive Modeling* (pp. 110–115).

Willingham, D. B., Salidis, J., & Gabrieli, J. D. E. (2002). Direct comparison of neural systems mediating conscious and unconscious skill learning. *Journal of Neurophysiology*, *88*(3), 1451–1460.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Year Two Project Report

Project: “Sparse sampling-based estimation of the retention of rarely used procedural knowledge.”

Principle Investigator: Prof. dr. H. van Rijn
Department of Experimental Psychology and Department of Psychometrics and Statistics
University of Groningen, The Netherlands

Grant number: FA9550-16-1-0043

Period of performance: November 2016 - November 2017

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Table of Contents

Summary ***3***

Introduction ***4***

Methods, Assumptions and Procedures ***5***

Results ***8***

Conclusions ***9***

References ***10***

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Summary

The main progress of 2017 was the completion of a study concerning the assessment of Extended Long-Term Retention of Script-Based Procedural Knowledge. Skills are typically learned to be applicable over much longer timespans, often expressed in years. Given that very frequent refresher sessions are impractical, many skills are only trained once every couple of months or even once per year (e.g., fire drills, Regulatory Reform (Fire Safety) Order 2005 (FSO); first aid, guidelines of the Dutch Red Cross; gun practice for police officers, Morrison, 2003), although it has been repeatedly argued that this frequency is too low (e.g., Morrison, 2003). To assess whether the Predictive Performance model (PPE; Jastrzembski & Gluck, 2009) is able to predict the retention of knowledge over extended periods of time, we ran an experiments with a much longer retention interval (4months). The target population were German students (at least 250 freshman German students enroll in the psychology program at the University of Groningen, The Netherlands) who have obtained their driver's license. Part of the requirements for obtaining a driver's license in Germany is CPR training. As most students will have obtained their driver's license at least a year before enrolling in the psychology program, these students are an excellent population to test long-term retention of skills. During the first session, we assessed their proficiency level after which training was provided to reestablish proficiency. We then tested their performance approximately four months after and reestablished proficiency, to evaluate whether the cognitive model can accurately predict performance at these extended time frames. In addition, declarative learning skills were obtained to assess the *relationship between the Retention of Declarative and Script-Based Procedural Knowledge*. Results were presented at the International Conference on Cognitive Modeling (Maass et al., 2017, see Appendix). Furthermore, due to a high drop-out rate a replication of this study was planned and the first data acquired at the end of 2017.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Introduction

Human memory entails both procedural and declarative knowledge. In lab settings, researchers often isolate either component and study it in controlled conditions. Most real-life behaviors and skills, however, draw on complex interactions of representations for both procedural and declarative components and are learned and retained over much longer time-frames than the typical lab-experiment. Research detailing the mechanics of procedural or declarative knowledge in isolation might therefore not directly transfer to situations in which both are involved. Hence it is important to study tasks in which both components interact, preferably complex real-life tasks spanning realistic time-frames, in controlled lab settings in order to improve our understanding of the relationship between these types of knowledge.

Cardiopulmonary resuscitation (CPR) is taught in the context of first aid training. It is a useful domain for studying the interaction of procedural and declarative aspects of memory because clear performance standards can be applied to quantifying learners' proficiency covering both components. CPR in terms of Basic Life Support, as laid out by the European Resuscitation Council (ERC) guidelines (Soar et al., 2015), includes an assessment of the so-called victim (check response, check breathing) and a series of steps (alert emergency services, hand positioning) before the actual compressions and rescue breaths are administered in a cycle of 30 compressions and two rescue breaths. Therefore, it combines declarative knowledge of which steps to perform and in which order as well as procedural knowledge of how to administer compressions and rescue breaths correctly. Using an instrumented Laerdal QCPR manikin, performance measures can be recorded and scored straightforwardly.

Crucially, CPR certification entails periodic re-training to ensure performance remains above criterion. For medical professionals, this is typically done every other year - an interval that is often considered suboptimal as performance might drop below the criterion during this interval (Stross, 1983). Furthermore, the American Heart Association (AHA) recognizes that large individual differences exist, which complicates the prescription of ideal methodology and frequency of training (Nolan et al., 2015). Consequently, acquisition and retention of CPR as a skill provides the ideal testbed for cognitive-psychological theories of learning and retention over realistic time-frames, with the potential of significant practical impact by optimizing training regimes.

The ultimate goal of any psychological theory is to predict behavior. The ability to predict future CPR performance based on behavior recorded at previous training moments is an especially interesting case because it presents the opportunity to predict, for each individual learner, when someone should return to be re-trained.

Collecting data in ecologically valid time frames, however, is extremely time consuming. However, a benefit of CPR is that subpopulations have been trained on this task before entering the lab. Specifically, part of the requirements to obtain a German driver's license is to demonstrate CPR performance above criterion. Therefore, German students with driver's licenses are a suitable population to test long-term retention of procedural and declarative memory because they had had CPR training in the past, typically had no retraining, and there will be natural variation in how long ago that training was completed.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Methods, Assumptions and Procedures

Participants

A total of 50 participants (37 female; 74%) with a median age of 20 (SD = 1.84; range = [18, 27]) were recruited from the participant pool at the University of Groningen. Of those, only 10 returned for the second session. (Due to this unexpectedly high drop-out rate, an additional study was planned for the winter 2017 / spring 2018. As data collection is still under way, we here report on this initial study.) Requirement for recruitment was the possession of a valid German driver's license. All participants gave informed consent and the Ethics Committee Psychology approved the study (ID: 17017-S-NE(a,b)).

Procedure and Stimuli

Participants completed two experimental sessions. A graphical summary of the paradigm is provided in Figure 1 and its components are described in more detail below.

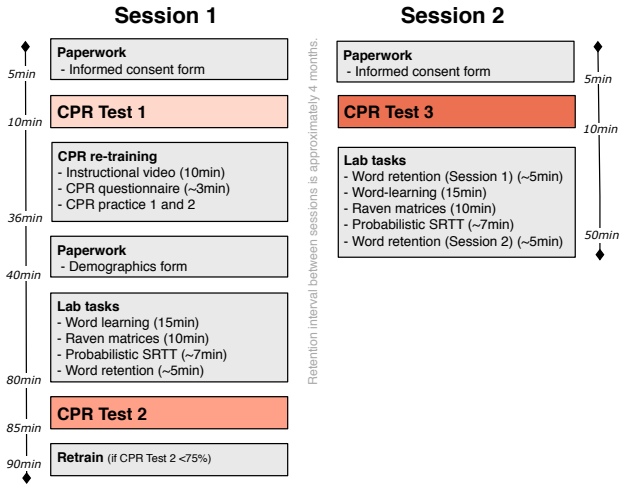


Figure 1. Overview of design and measurement moments.

CPR Test 1 At the beginning of each session, participants signed the informed consent forms. In the first session participants then entered the experimentation room where a Laerdal Resusci Anne QCPR manikin was lying on the ground. Participants read the following instructions: "*You volunteered for community service to help elderly neighbors with chores in their homes. When you enter the house of Mr. Johnson, you find him on the living room floor. There are no signs of bleeding or open wounds and no one else is in the house. Based on your first aid training, take the steps necessary in this situation on the manikin to* ***assess*** *and* ***react upon*** *the situation*."

This scenario was chosen to sketch a hypothetical scenario that required them to perform CPR on the manikin. Specifically, it stated that they are alone, making it necessary to act and not call additional first aid helpers. Any bleeding and wounds were excluded to ensure that participants do not waste time to check for such. Participants were then asked to perform the necessary actions required according to the ERC guidelines (Soar et al. 2015). This means, participants were supposed to alternate between 30 compressions and two rescue breaths. Participants were stopped after administering four cycles of compressions and rescue breaths (i.e., 30-2, 30-2, etc.) to avoid fatigue. We refer to this procedure (i.e., initial steps followed by four rounds of 30-2) as one run-through of CPR.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

The initial steps were scored using a pen and paper checklist and the administration of compressions and rescue breaths were recorded on the manikin's SimPad. Performance scores for chest compressions and rescue breaths were based on Laerdal's proprietary scoring algorithm using the European guidelines (ranging from 0 to 100%; a score of 75% or higher is considered "proficient"). The SimPad outputs overall scores for compressions, ventilations, and a combined score. The data presented here will focus on the combined scores. This procedure was identical for CPR Tests in all sessions.

CPR Re-Training After the initial assessment, participants were re-trained. First, participants watched a short instructional video (see https://osf.io/9er6g/) demonstrating the initial steps, as well as instructions on how to correctly apply chest compressions and rescue breaths. This video was specifically made for this research project.

Subsequently, participants had the opportunity to practice compressions on the manikin with its live feedback option enabled for one minute. That is, for each compression participants could track their depth and frequency and adjust if necessary. Then, participants practiced giving rescue breaths until the live feedback indicated that two correct breaths had been given in a row.

CPR Practice 1 and 2 Participants were instructed to "*Perform the complete procedure you saw in the video, with four rounds of compressions and rescue breaths*" twice while their performance was scored.

After the run-throughs of CPR, participants completed questionnaires to gather demographic information, the date their driver's license was issued, and the approximate number of months between obtaining their license and completing their CPR training. The time between the mandatory training and obtaining the driver's license ranged from 1 to 60 months (mean = 9.92 and SD = 12.71).

Word Learning Study Participants spent 15 minutes studying 35 Swahili-English word pairs using the adaptive method outlined in the Introduction. The word pairs were taken from Nelson and Dunlosky (1994) and items did not repeat across sessions.

Raven Matrices Participants performed a shortened version of Raven's advanced progressive matrices (Raven, 2000). Instead of seeing all 36 matrices without a time limit, participants had 10 minutes to solve half of them (as in Foster et al., 2015). Participants saw the 18 odd-numbered items in the first session and the 18 even-numbered items in the second session.

Serial Reaction Time Task Participants performed 768 trials (8 blocks of 96 trials) of a the SRTT task.

CPR Test 2 Following the computerized tasks, participants were asked to complete another complete run-through of CPR. If the score of this test was below 75%, participants were re-trained until they reached criterion.

Word Learning Test The first session ended with a retention test of the Swahili words the participants learned during Word Learning Study. Participants were asked to fill in all of the English translations they remembered when presented with the Swahili word. In the second session participants completed the same word retention test, in addition to the word retention test for the stimuli set of Session 2.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Predicting Future CPR Performance All participants had completed first aid before obtaining their driver's license. Although the approximate date of this initial training is known, no performance measures are available. Hence, all participants' prior performance was assumed to be at passing rate (i.e., 75%). The scores of CPR Test 1 and 2 in the first session and the time point of the initial training with the corresponding score of 75% were used to make a prediction for the second session using PPE (Jastrzembski & Gluck, 2009).

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Results

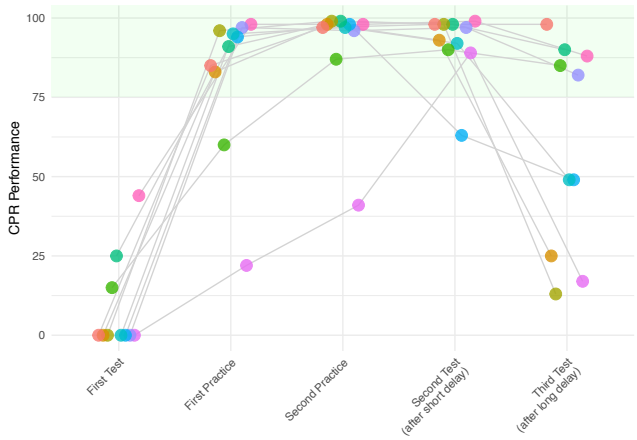


Figure 2. CPR Training and Retention scores of the subjects that completed both sessions.



Figure 3. Measures on Lab tasks (for subjects that completed both sessions)

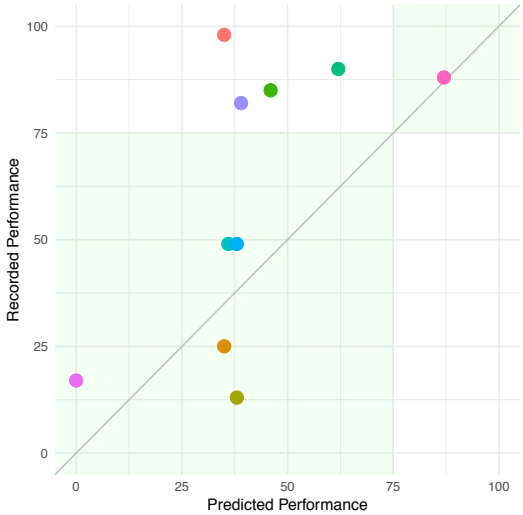


Figure 4. Model Predictions and Recorded performance of subjects that completed both sessions.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Conclusions

The experiment demonstrated a successful translation of a real-world task into a controlled lab setting. Promising results were found of predicting the performance on CPR over a time period of four months. Yet, due to a high-dropout rate a replication of this study is necessary which has been planned for 2017/2018 (currently on going). Furthermore, this experiment showed that CPR performance is correlated with the assessment of declarative learning (better than "IQ" scores). However, inconclusive SRTT results preclude comparisons with CPR data.

Already at this stage, this study has important real-world implications, as it demonstrates that a single CPR training session (i.e., training participants once to achieve successful performce), even in young and highly-educated individuals, does not achieve reasonable CPR retention. In other words, the strong decay of performance between driver-license related CPR training and the initial tests when participants entered our study indicates that a single CPR training is insufficient for long term retention. However, the noticeable increases between initial test and the test after watching a CPR-training video indicate that this population can quickly, easily, and thus inexpensively be brought back at the required level of performance. And although due to the small number of participants the results of the delayed tests should be interpreted with caution, the reasonable retention rates indicate that personalized training schedules might provide a useful method for training complex skills such as CPR.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

References

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, 43(2), 226-236.

Jastrzembski, T. S., & Gluck, K. A. (2009). *A formal comparison of model variants for performance prediction*. AIR FORCE RESEARCH LAB MESA AZ HUMAN PERFORMANCE WING (711TH) HUMAN EFFECTIVENESS DIRECTORATE/WARFIGHTER READINESS RESEARCH DIVISION.

Maass, S., Sense, F., Walsh, M., Gluck, K., & Van Rijn, H. (2017). Cognitive Modeling of Cardiopulmonary Resuscitation Knowledge and Skill Spanning Months to Years (abstract). In M. K. van Vugt, A. P. Banks, & W. G. Kennedy (Eds.), *Proceedings of the 15th International Conference on Cognitive Modeling* (pp. 228–229). Coventry, UK: University of Warwick.

Morrison, L. J., Neumar, R. W., Zimmerman, J. L., Link, M. S., Newby, L. K., McMullan, P. W., ... & Edelson, D. P. (2013). Strategies for improving survival after in-hospital cardiac arrest in the United States: 2013 consensus recommendations: a consensus statement from the American Heart Association. *Circulation*, *127*(14), 1538-1563.

Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents*. Memory, 2*(3), 325-335.

Nolan, J. P., Hazinski, M. F., Aickin, R., Bhanji, F., Billi, J. E., Callaway, C. W., ... & Gent, L. M. (2015). Part 1: Executive Summary: 2015 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science with Treatment Recommendations. *Resuscitation*, 95, e1-e31.

Soar, J., Nolan, J. P., Böttiger, B. W., Perkins, G. D., Lott, C., Carli, P., ... & Sunde, K. (2015). European Resuscitation Council guidelines for resuscitation 2015. *Resuscitation*, 95, 100-147.

Stross, J.K. (1983). Maintaining Competency in Advanced Cardiac Life Support Skills. *Journal of the American Medical Association*, 249(24), 3339–3341.

Appendix

Poster ICCM 2017: Maass, S., Sense, F., Walsh, M., Gluck, K., & Van Rijn, H. (2017). Cognitive Modeling of Cardiopulmonary Resuscitation Knowledge and Skill Spanning Months to Years.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Predicting Performance in a Complex Real-Life Task: Learning and Forgetting CPR

Sarah Maass* (s.c.maass@rug.nl) & Florian Sense* (f.sense@rug.nl)
Department of Experimental Psychology, University of Groningen
Groningen, The Netherlands
** Shared first authorship, alphabetical order*

Michael Krusmark (michael.krusmark.ctr@us.af.mil)
L3 Technologies at Air Force Research Laboratory, Wright-Patterson Air Force Base
Dayton, Ohio, USA

Kevin Gluck (kevin.gluck@us.af.mil)
Air Force Research Laboratory, Wright-Patterson Air Force Base
Dayton, OH, USA

Hedderik van Rijn (d.h.van.rijn@rug.nl)
Department of Experimental Psychology, University of Groningen
Groningen, The Netherlands

Abstract

In this study we investigate learning and memory in the real world task domain of cardiopulmonary resuscitation (CPR), a basic lifesaving skill that taps into a combination of procedural and declarative learning. Participants all had previous CPR training months to years earlier, creating an unusual opportunity to assess longer-term retention and model parameterization. In the first experimental session, participants' current CPR skill level was assessed, then they were re-trained and performed several bouts of CPR. The goal was to predict performance levels after 8 and 16 weeks of this training, using the Predictive Performance Equation (PPE; Walsh, Gluck, Gunzelmann, Jastrzembski, & Krusmark, in press). With learning and retention periods spanning months to years, it allowed us to assess the predictive accuracy of PPE over an ecologically relevant timeframe.

Keywords: skill acquisition; retention; learning; memory; CPR

Introduction

Human memory entails both procedural and declarative knowledge. In lab settings, researchers often isolate either component and study it in controlled conditions. Most real-life behaviors and skills, however, draw on complex interactions of representations for both procedural and declarative components and are learned and retained over much longer time-frames than the typical lab-experiment. Research detailing the mechanics of procedural or declarative knowledge in isolation might therefore not directly transfer to situations in which both are involved. Hence it is important to study tasks in which both components interact, preferably complex real-life tasks spanning realistic time-frames, in controlled lab settings in order to improve our understanding of the relationship between these types of knowledge.

Cardiopulmonary resuscitation (CPR) is taught in the context of first aid training. It is a useful domain for studying the interaction of procedural and declarative aspects of memory because clear performance standards can be applied to quantifying learners' proficiency covering both components. CPR in terms of Basic Life Support, as laid out by the European Resuscitation Council (ERC) guidelines (Soar et al., 2015), includes an assessment of the so-called victim (check response, check breathing) and a series of steps (alert emergency services, hand positioning) before the actual compressions and rescue breaths are administered in a cycle of 30 compressions and two rescue breaths. Therefore, it combines declarative knowledge of which steps to perform and in which order as well as procedural knowledge of how to administer compressions and rescue breaths correctly. Using an instrumented Laerdal QCPR manikin, performance measures can be recorded and scored straightforwardly.

Crucially, CPR certification entails periodic re-training to ensure performance remains above criterion. For medical professionals, this is typically done every other year - an interval that is often considered suboptimal as performance might drop below the criterion during this interval (Stross, 1983). Furthermore, the American Heart Association (AHA) recognizes that large individual differences exist, which complicates the prescription of ideal methodology and frequency of training (Nolan et al., 2015). Consequently, acquisition and retention of CPR as a skill provides the ideal testbed for cognitive-psychological theories of learning and retention over realistic time-frames, with the potential of significant practical impact by optimizing training regimes.

The ultimate goal of any psychological theory is to predict behavior. The ability to predict future CPR performance based on behavior recorded at previous training moments is an especially interesting case because it presents the opportunity to predict, for each individual learner, when someone should return to be re-trained.

Collecting data in ecologically valid time frames, however, is extremely time consuming. However, a benefit of CPR is that subpopulations have been trained on this task before

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

entering the lab. Specifically, part of the requirements to obtain a German driver's license is to demonstrate CPR performance above criterion. Therefore, German students with driver's licenses are an suitable population to test long-term retention of procedural and declarative memory because they had had CPR training in the past, typically had no retraining, and there will be natural variation in how long ago that training was completed.

Mathematical models of skill acquisition and retention can help describe fluctuations in CPR performance over time. One such model, the Predictive Performance Equation (PPE), formalizes the temporal dynamics in a series of mathematical equations (Walsh, Gluck, Gunzelmann, Jastrzembski, & Krusmark, in press). First, performance increases with the amount of practice (*N*). This is implemented as a power law of learning, with the *learning rate* fixed at 0.1 on the basis of prior empirical evidence and model fits. Second, performance drops as a function of elapsed time among practice events (*T*). This is implemented as a power function of forgetting, with the decay rate determined by the function expressed in Equation 4, below. In PPE the effects of learning and retention on activation is multiplicative, such that:

$$activation = N^{learning\ rate} \cdot T^{-decay\ rate} \quad (1)$$

Third, PPE captures the spacing effect, such that retention is better and more stable when practice is distributed over time. This is implemented in the forgetting function through *T* and the *decay rate*. *T* is computed as the sum of the weighted age of each practice event,

$$T = \sum_{i=1}^{n-1} w_i \cdot t_i, \quad (2)$$

where the weight is an exponential decay function of time,

$$w_i = t_i^{-x} \sum_{j=1}^{n-1} \frac{1}{t_j^{-x}}. \quad (3)$$

Thus, *T* weighs practice repetitions so that more recent events carry more weight, and the variable *x*, which is fixed at 0.6, controls the degree of the weighting.

The *decay rate* is computed as a function of the complete history of lags between successive practice opportunities (lag_i):

$$decay\ rate = decay\ intercept + decay\ slope \cdot average\left(\frac{1}{log(lag_i)}\right) \quad (4)$$

Finally, in PPE performance is computed as a logistic function of activation:

$$Performance = \frac{1}{1 + exp\left(\frac{threshold - activation}{scalar}\right)} \quad (5)$$

PPE parameters are estimated separately for individuals based on their performance histories. For each individual, we compute the optimal values of the *decay intercept*, *decay slope*, *threshold*, and *scalar* that minimizes the mean squared deviation between model and human performance. These parameter values are then used to generate out of sample predictions of performance at future points in time.

An increasingly common modeling practice in sparse data environments is to seed parameter values with priors. There has been some previous exploration of this approach in the context of PPE (Collins, Gluck, Walsh, & Krusmark, 2017). Thus, to enhance the quality of our predictions of CPR performance on Session 2 in the current study, we utilized priors generated from a longer-term CPR study that is currently underway (Jastrzembski, et al, 2017). In initial analyses from this study, CPR compression data from four sessions separated by either one day, one week, one month, or three months were used to calibrate the model and generate model parameters for temporal predictions at either three or six months in the future. These model parameters were utilized in the current study to inform parameters generated during the model fitting process.

The PPE is used to model and predict changes in performance between sessions based on the time that elapsed. Using similar mechanics/equations, one can also optimize repetition schedules *within* individual learning sessions (Van Rijn, Van Woudenberg, Van Maanen, 2009; Sense & Van Rijn, submitted). Earlier work has shown that an ACT-R based cognitive model can use response accuracy and latency on a trial-by-trial basis to predict when each studied item is likely to be forgotten and ensure rehearsal before that moment. This improves retention of the studied facts (Sense & Van Rijn, submitted) and allows the estimation of a learner's rate of forgetting (Sense, Behrens, Meijer, Van Rijn, 2016). This adaptive fact-learning model estimates a decay term for each item per learner and by averaging across the item-level parameter values, a rate of forgetting can be extracted for each participant, indicating how quickly they forget studied material. Here, we will test whether this rate of forgetting is related to their decrease in CPR performance measured during the course of the current study.

In addition, we administered Raven matrices, a task frequently used as a rough estimate of general cognitive ability (Raven, 2000). If performance on the Raven matrices task is related to someone's ability to acquire and retain CPR knowledge, such a score could provide a potential covariate to refine the predictive model.

The current study was devised to assess the accuracy of personalized performance predictions. The ERC's guidelines (Soar et al., 2015) state that "*The intervals for retraining will differ according to the characteristics of the participants*". The availability of predictive models that take an individual's performance profile over time into account permits such personalized predictions. Ideally, this would make re-trainings more efficient and reduce the interval during which

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

a medical professional might perform below criterion. The viability of personalized refresher schedules crucially depends on the accuracy of the predictions: Requiring people to retrain too early is a waste of resources but requiring them too late can cost lives.

Methods

Participants

A total of 50 participants (37 female; 74%) with a median age of 20 (SD = 1.84; range = [18, 27]) were recruited from the participant pool at the University of Groningen. Of those, 10 dropped out after the first session. Requirement for recruitment was the possession of a valid German driver's license. Two participants indicated minimal familiarity with Swahili, which is relevant to the paired associates word learning task in this study. All participants gave informed consent and the Ethics Committee Psychology approved the study (ID: 17017-S-NE(a,b)).

Procedure and Stimuli

Participants completed three experimental sessions. CPR performance was assessed in all sessions and computerized tasks were administered in Session 1 and 2. A graphical summary of the paradigm is provided in Figure 1 and its components are described in more detail below.

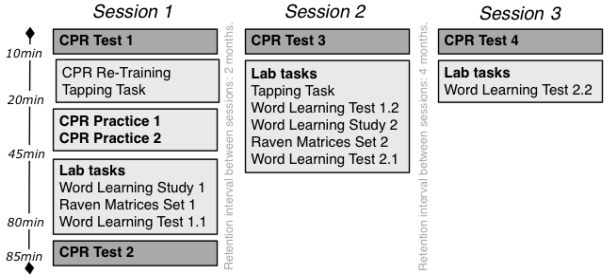


Figure 1. Overview of design and measurement moments.

CPR Test 1 At the beginning of each session, participants signed the informed consent forms. In the first session participants then entered the experimentation room where a Laerdal Resusci Anne QCPR manikin was lying on the ground. Participants read the following instructions: "*You volunteered for community service to help elderly neighbors with chores in their homes. When you enter the house of Mr. Johnson, you find him on the living room floor. There are no signs of bleeding or open wounds and no one else is in the house. Based on your first aid training, take the steps necessary in this situation on the manikin to* ***assess*** *and* ***react upon*** *the situation*."

This scenario was chosen to sketch a hypothetical scenario that required them to perform CPR on the manikin. Specifically, it stated that they are alone, making it necessary to act and not call additional first aid helpers. Any bleeding and wounds were excluded to ensure that participants do not waste time to check for such. Participants were then asked to perform the necessary actions required according to the ERC guidelines (Soar et al. 2015). This means, participants were supposed to alternate between 30 compressions and two rescue breaths. Participants were stopped after administering four cycles of compressions and rescue breaths (i.e., 30-2, 30-2, etc.) to avoid fatigue. We refer to this procedure (i.e., initial steps followed by four rounds of 30-2) as one run-through of CPR.

The initial steps were scored using a pen and paper checklist and the administration of compressions and rescue breaths were recorded on the manikin's SimPad. Performance scores for chest compressions and rescue breaths were based on Laerdal's proprietary scoring algorithm using the European guidelines (ranging from 0 to 100%; a score of 75% or higher is considered "proficient"). The SimPad outputs overall scores for compressions, ventilations, and a combined score. The data presented here will focus on the combined scores. This procedure was identical for CPR Tests in all sessions.

CPR Re-Training After the initial assessment, participants were re-trained. First, participants watched a short instructional video (see https://osf.io/9er6g/) demonstrating the initial steps, as well as instructions on how to correctly apply chest compressions and rescue breaths. This video was specifically made for this research project.

Subsequently, participants had the opportunity to practice compressions on the manikin with its live feedback option enabled for one minute. That is, for each compression participants could track their depth and frequency and adjust if necessary. Then, participants practiced giving rescue breaths until the live feedback indicated that two correct breaths had been given in a row.

Tapping Task Participants completed five minutes of the Continuation and Synchronization Tapping Task (Semjen, Schulze, & Vorberg,2000). Due to space limitations, results from this task will not be discussed in this manuscript.

CPR Practice 1 and 2 Participants were instructed to "*Perform the complete procedure you saw in the video, with four rounds of compressions and rescue breaths*" twice while their performance was scored.

After the run-throughs of CPR, participants completed questionnaires to gather demographic information, the date their driver's license was issued, and the approximate number of months between obtaining their license and completing their CPR training. The time between the mandatory training and obtaining the driver's license ranged from 1 to 60 months (mean = 9.92 and SD = 12.71).

Word Learning Study Participants spent 15 minutes studying 35 Swahili-English word pairs using the adaptive method outlined in the Introduction. The word pairs were taken from Nelson and Dunlosky (1994) and items did not repeat across sessions.

Raven Matrices Participants performed a shortened version of Raven's advanced progressive matrices (Raven, 2000). Instead of seeing all 36 matrices without a time limit, participants had 10 minutes to solve half of them (as in Foster

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

et al., 2015). Participants saw the 18 odd-numbered items in the first session and the 18 even-numbered items in the second session.

CPR Test 2 Following the computerized tasks, participants were asked to complete another complete run-through of CPR. If the score of this test was below 75%, participants were re-trained until they reached criterion.

Word Learning Test The first session ended with a retention test of the Swahili words the participants learned during Word Learning Study (Test 1.1). Participants were asked to fill in all of the English translations they remembered when presented with the Swahili word. In the second session participants completed the same word retention test (Test 1.2), in addition to the word retention test for the stimuli set of Session 2 (Test 2.1), which will be tested again in Session 3 (Test 2.2).

Predicting Future CPR Performance

All participants had completed first aid before obtaining their driver's license. Although the approximate date of this initial training is known, no performance measures are available. Hence, all participants' prior performance was assumed to be at passing rate (i.e., 75%). The scores of CPR Test 1 and 2 in the first session and the time point of the initial training with the corresponding score of 75% were used to make a prediction for the second session using PPE (see Introduction). These predictions can be compared with the actual performance on CPR Test 3 (see Figure 3). Importantly, we will use the data in Figure 2 to predict performance during CPR Test 4, scheduled for June 2018.

Results

Here we provide descriptive statistics and visualizations of the data through the end of Session 2. Data collection for the third session is scheduled for May/June 2018 and once the complete dataset is available, specific hypotheses and ideas will be tested more formally. In the current context, we will focus on the following: CPR performance at the different test moments, PPE predictions for Session 2 and how they match the collected data, and the correlations between the lab tasks and the CPR components.

The overall CPR performance assessed in the first two sessions is shown in Figure 2. Shown are the scores exported from Laerdal's proprietary software, which combines the compression and rescue breath performance into a single score. The area corresponding to "proficient" performance is shaded green. Measurements from a single participant are grouped by grey lines and color-coded. Overlaid are "violins" to summarize the distribution of scores at each time point.

On the initial test at the beginning of the first session, only two participants demonstrate proficient performance, while many score below 25%. The CPR Re-Training administered between CPR Test 1 and CPR Practice 1 results in a marked increase in performance, making the majority of participants reach criterion. Testing for a difference between those two scores with a paired Bayesian t-test yields an decisive Bayes factor of 2.4×10^{17} in favor of a difference. The second practice marks a further increase in overall performance and the vast majority of participants retain above-threshold performance until CPR Test 2 at the end of the first session. In the eight-week interval until the second session, and performance decreases ($BF_{H1} = 9.95$) but many participants still exhibit near-ceiling performance.

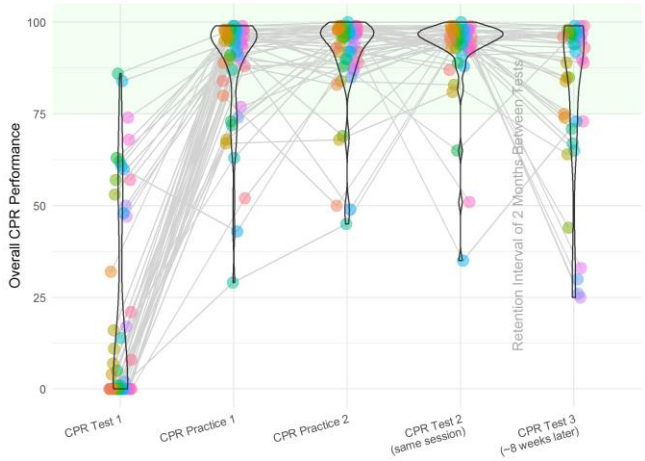


Figure 2. Overall CPR performance at the different measurement moments highlighted in Figure 1.

The recorded performance at the beginning of Session 2 (CPR Test 3 in Figure 2) can be contrasted with the predictions generated using PPE. A graphical summary is provided in Figure 3. The two quadrants shaded in green contain participants for whom predicted and actual performance match relative to the 75% proficiency cut-off. The yellow and red quadrants contain participants for whom performance was under- or over-predicted, respectively. Predictions were made for all 50 participants but actual performance scores are only available for the 39 participants how participated in Session 2. Of those, the model predicted that 11 participants would perform above threshold, compared to an empirical total of 27. The model predicted that 28 participants would perform below threshold, compared to an empirical total of 12. The predictions and actual performance are positively correlated ($r = 0.592$), with a root-mean-square error at 32.1. The high error is caused by the cluster in the yellow region, who perform near-perfect while the model predicted scores between 25 and 75%.

The two relevant lab tasks - Raven Matrices and Word Learning - were administered twice each (cf. Figure 1). For the Raven Matrices task, participants saw 16.1 out of 18 problems in the first, and 16.4 in the second session. Of those problems, they solved an average of 9.3 in both sessions.

For the Word Learning task, participants studied an average of 21.3 and 22.2 items (out of 35) in the two sessions, respectively. Each participant's rates of forgetting (one per session) were estimated by averaging across the final value of the estimated parameter associated with each item they encountered at least three times.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

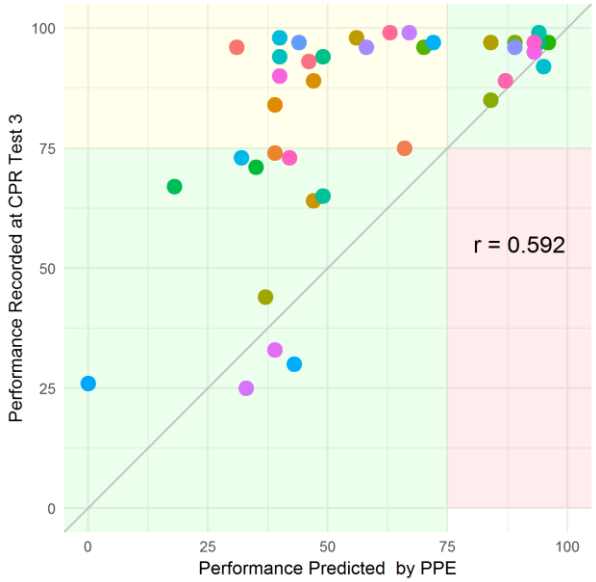


Figure 3. Comparing predicted performance with recorded performance for participants that returned for Session 2.

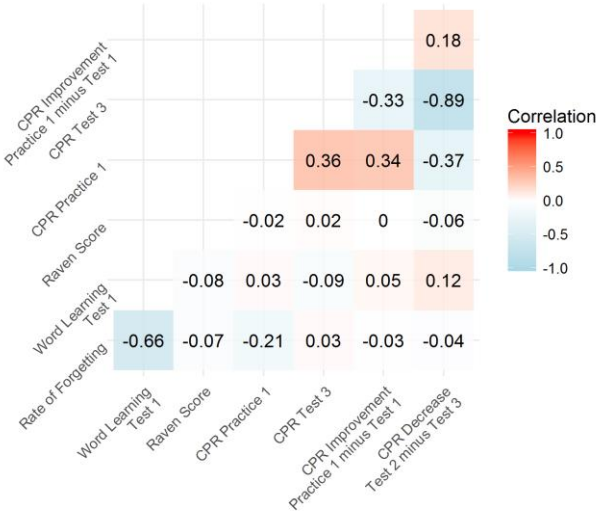


Figure 4. Color-coded correlation matrix for a subset of the relevant outcome measures.

A more detailed overview of the correlations between some of the recorded measures is provided in Figure 4. Given the limited space in the current format, we will focus here on a selection of measures and their correlations. Specifically, we will focus on performance measures from the lab tasks from the first session: the rate of forgetting estimated during Word Learning, the score on the recall test of those same words, as well as the number of Raven matrices solved. These are correlated with four data points that summarize someone's CPR performance: scores on the first practice (i.e., after seeing the instructions), CPR performance after a 2-month delay (i.e., CPR Test 3), as well as the two difference scores corresponding to the Bayes factors reported above. The first captures CPR Improvement from the initial assessment to performance after having received instructions, and the second captures CPR Decrease between the last assessment in first session (CPR Test 2) and the first assessment in the second session (CPR Test 3, see Figure 2).

Note that a sizeable proportion of the correlation coefficients are very small (i.e., < 0.10). Specifically, the Raven scores do not seem to correlate with any of the other measures, neither do the scores on the delayed recall test from the Word Learning task. The estimated rate for forgetting correlates highly with performance on the recall test and the negative correlation indicates that a faster rate of forgetting results in lower test scores (as expected).

The correlation coefficients between the different CPR scores, on the other hand, are more substantial. Most striking is the strong negative relationship (r = -0.89) between someone's performance in Session 2 (CPR Test 3) and the decrease in performance between the first two sessions, which makes sense because someone scoring low on CPR Test 3 must have decreased in performance (also see Figure 2). Similarly, participants whose CPR performance improves tend to score higher on the first practice (r = 0.34) and subsequently also tend to score higher in the second session (CPR Test 3; r = 0.36).

Interestingly, the correlation between the estimated rate of forgetting and the decrease in performance between the first two sessions is almost zero (r = -0.04), suggesting that the rate with which a participant is estimated to forget factual knowledge (within a learning session) is not related to their decrease in CPR performance (between sessions).

Discussion

The main focus of the current work is an attempt to predict future cardiopulmonary resuscitation (CPR) performance using previous CPR performance. Additional lab tasks were administered to explore whether they are related to CPR performance and could be used to further improve performance of future CPR performance. After completion of the first session, predictions have been made for CPR performance in the second session, using performance in Session 1 along with approximated performance and the time between initial training and Session 1. Figure 3 contrasts predicted and recorded performance for Session 2. These data will be incorporated in PPE to yield predictions for Session 3, which is to be conducted in May/June 2018.

As can be seen in Figure 3, the predictions are not perfect. In fact, there is a large group of participants that were predicted to perform below the proficiency threshold but performed well above it (i.e., the yellow region). Importantly, however, there are no participants in the red region of the figure: *No participant was falsely predicted to be proficient*.

From a model-building perspective, a model making unbiased predictions would be preferred over a biased model. With the focus on a real-life task that is intimately connected to basic lifesaving skills, however, there is an argument to be made for erring on the conservative side, as the costs of performing below criterion can be enormous. Once we have

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

collected the data for Session 3, we can address whether the accuracy (e.g., RMSE) of the predictions will increase (which we expect) and whether the (preferred) conservative bias will persist or whether prediction errors will fall in the red region of Figure 3 as well. The prediction is made easier by the availability of more data but more challenging as the retention interval is twice as long. A practical question that will also remain open until more data are collected is the number of measurement moments that are sufficient to make predictions that are acceptably reliable in real-world contexts. Only then can schedules for personalized refresher trainings be relevant in practice.

The lab tasks were chosen to assess a general measure of cognitive ability (Raven matrices) and a measure of declarative learning (Word Learning). Both tasks are more constrained than the CPR performance measures that capture someone's skill. The lack of correlations between the lab tasks and the various CPR scores (see Figure 4) suggests that performance on the simpler tasks does not predict performance on the more complex tasks. More detailed analyses will be conducted once the complete dataset is available. However, the lack of correlations between the components reported here suggests that predicting future CPR performance is not improved by incorporating scores derived from these traditional lab-tasks.

The ERC's report (Soar et al., 2015) states that CPR performance is known to deteriorate within months of training and, therefore, even annual re-trainings might not be frequent enough for some people. Since CPR training can be time consuming and optimal training intervals are currently unknown, they suggest that frequent "low dose" training using video instructions and hands-on practice can be as effective as instructor-led courses (Nolan et al., 2015). The work presented here confirms that there is a swift improvement in performance after such CPR Re-Training (cf. CPR Test 1 vs CPR Practice 1 in Figure 2)

In summary, we report an experimental setup in which the learning and forgetting of CPR is assessed over ecologically relevant timeframes. We test a mathematical model's ability to predict future CPR performance using very sparse data. A first wave of predictions is presented here and the second will be presented at the conference. We anticipate predictive accuracy to improve for Session 3 and will be able to evaluate the feasibility of personalized schedules for refresher courses.

Acknowledgments

This work was supported by EOARD grant #11926121.

References

Collins, M. G., Gluck, K. A., Walsh, M. M., & Krusmark, M. A. (2017). Using prior data to inform initial performance predictions on individual students, In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*.

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, 43(2), 226-236.

Jastrzembski, T., Walsh, M. M., Krusmark, M., Kardong-Edgren, S., Oermann, M., Dufour, K., Millwater, T., Gluck, K. A., Gunzelmann, G., Harris, J., & Stefanidis, D. (2017). Personalizing training to acquire and sustain competence through use of a cognitive model. In D. D. Schmorrow & C. M. Fidopiastis, *Augmented cognition. Enhancing cognition and behavior in complex environments* (pp. 148-161). Switzerland: Springer International Publishing AG.

Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory, 2*(3), 325-335.

Nolan, J. P., Hazinski, M. F., Aickin, R., Bhanji, F., Billi, J. E., Callaway, C. W., ... & Gent, L. M. (2015). Part 1: Executive Summary: 2015 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science with Treatment Recommendations. *Resuscitation*, 95, e1-e31.

Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cognitive Psychology*, 41(1), 1-48.

Semjen, A., Schulze, H. H., & Vorberg, D. (2000). Timing precision in continuation and synchronization tapping. *Psychological Research*, 63(2), 137-147.

Sense, F., Behrens, F., Meijer, R. R., & van Rijn, H. (2016). An Individual's Rate of Forgetting Is Stable Over Time but Differs Across Materials. *Topics in Cognitive Science*, 8: 305–321.

Sense, F., & Van Rijn, H. (submitted). Optimizing Within-Session Repetition Schedules with a Response-Latency-Based Cognitive Model Improves Retention.

Soar, J., Nolan, J. P., Böttiger, B. W., Perkins, G. D., Lott, C., Carli, P., ... & Sunde, K. (2015). European Resuscitation Council guidelines for resuscitation 2015. *Resuscitation*, 95, 100-147.

Stross, J.K. (1983). Maintaining Competency in Advanced Cardiac Life Support Skills. *Journal of the American Medical Association*, 249(24), 3339–3341.

van Rijn, H., van Maanen, L., & van Woudenberg, M. (2009). Passing the Test: Improving Learning Gains by Balancing Spacing and Testing Effects. In A. Howes, D. Peebles, & R. P. Cooper (Eds.), Proceedings of the 9th International Conference on Cognitive Modeling (pp. 110–115). Manchester, UK.

van Rijn, H., van Maanen, L., & van Woudenberg, M. (2009). Passing the Test: Improving Learning Gains by Balancing Spacing and Testing Effects. In A. Howes, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 9th International Conference on Cognitive Modeling* (pp. 110–115). Manchester, UK.

Walsh, M. M., Gluck, K. A., Gunzelmann, G., Jastrzembski, T., & Krusmark, M. (in press). Evaluating the theoretical adequacy and applied potential of computational models of the spacing effect. *Cognitive Science*.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Cognitive Modeling of Cardiopulmonary Resuscitation Knowledge and Skill Spanning Months to Years

Sarah **Maass**[1], Florian **Sense**[1], Matthew **Walsh**[2], Kevin **Gluck**[3], Hedderik **van Rijn**[1]

[1] Experimental Psychology, University of Groningen, NL; [2] TiER 1 Performance Solutions, Pittsburgh, PA, USA; [3] Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, USA

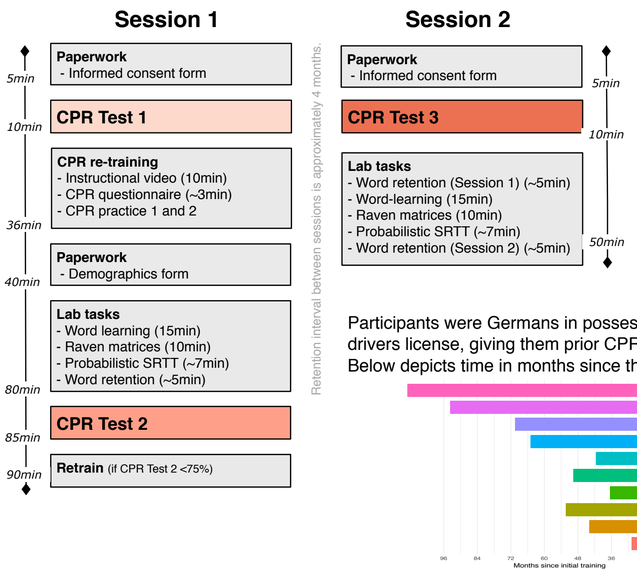
Aim: study learning and memory in a real-world task and predict performance using the Predictive Performance Equation

Cardiopulmonary Resuscitation (CPR) is a basic lifesaving skill that combines procedural and script-based learning.
Predicting CPR performance has important societal consequences, such as prescribe personalized refresher training intervals.
CPR performance was assessed along with standard lab tasks to explore relationships with standardized procedural and declarative tasks :
Procedural Learning: ***Serial Reaction Time Task (SRTT)*** , Declarative Learning: ***Adaptive fact-learning system,*** IQ: ***Raven Matrices***

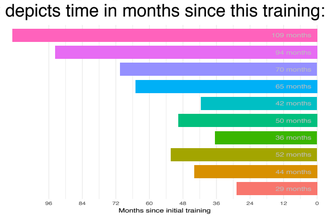


CPR assessment with modern, instrumented QCPR manikin

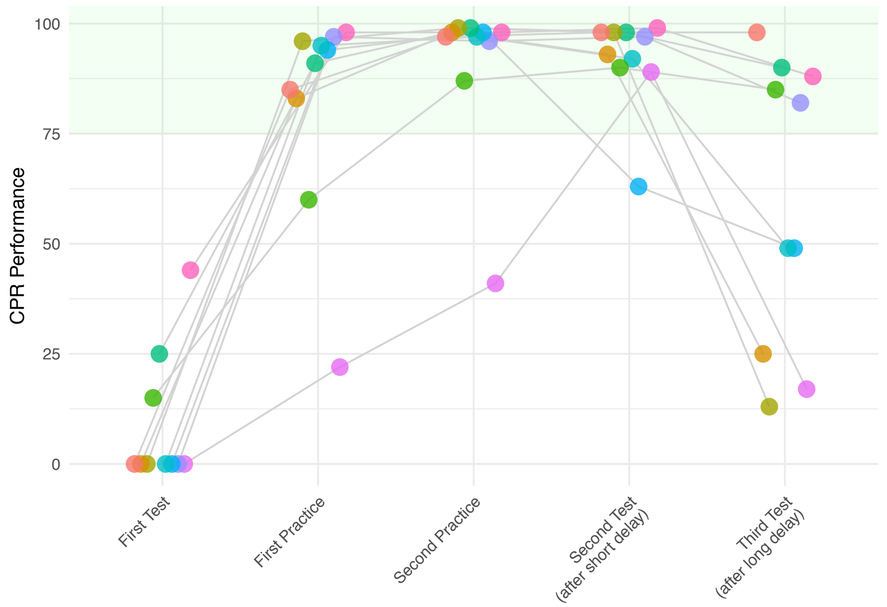
Experimental Set-up



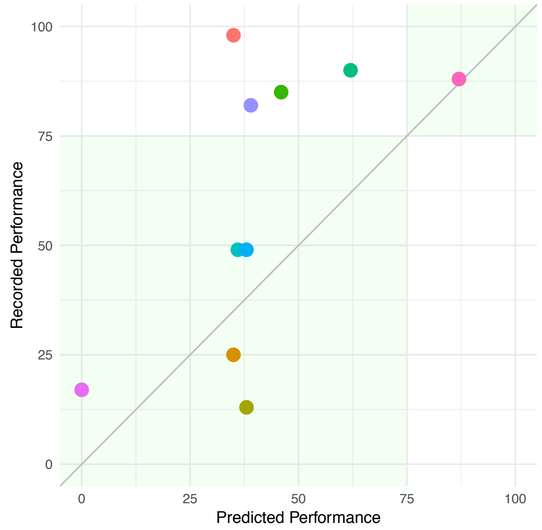
Participants were Germans in possession of a drivers license, giving them prior CPR training. Below depicts time in months since this training:



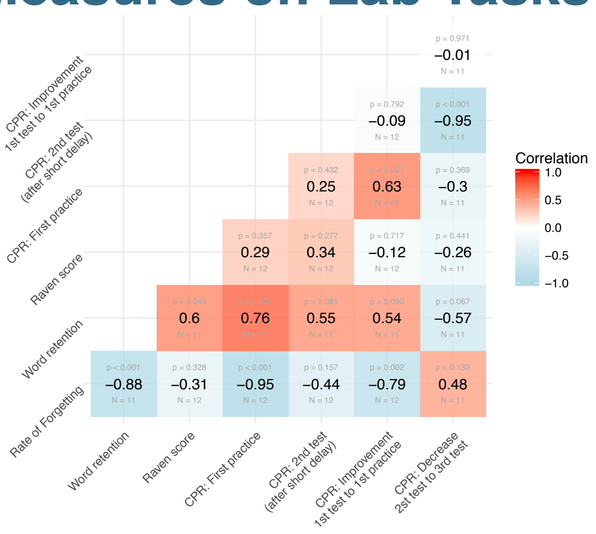
CPR Training & Retention



Model Prediction: PPE



Measures on Lab Tasks



Conclusions

- ✓ Successful translation of a real-world task into a controlled lab setting
- ✓ Promising results of predicting performance on CPR over ~ 4 months
- ✓ CPR performance is correlated with the assessment of declarative learning (better than "IQ" scores)
- ✗ Inconclusive SRTT results preclude comparisons with CPR data

Real-world Implications:

- Strong decay of performance prior to First Test is indicative of a single CPR training being insufficient for long term retention
- Results of Third Test shows importance of personalized training schedules

The **Predictive Performance Equation** (PPE; Jastrzembski & Gluck, 2009; Walsh, Gluck, Gunzelmann, Jastrzembski, & Krusmark, in revision) predicts performance based on:

1. Amount of practice
2. Elapsed time since practice
3. Temporal spacing of practice

$$\text{Performance} = \text{Practice}^{c} \times \text{Elapsed Time}^{-d}$$

$$c = \text{Learning rate}$$

$$d = \text{Decay intercept} + \text{Spacing slope} \times \text{mean}\left(\frac{1}{\log(lags)}\right)$$

Including more space between practice repetitions reduces decay rate, thereby increasing retention.

contact: s.c.maass@rug.nl

Acknowledgements: This work was supported by EOARD grant 11926121.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Year Three Project Report

Project: “Sparse sampling-based estimation of the retention of rarely used procedural knowledge.”

Principle Investigator: Prof. dr. H. van Rijn
Department of Experimental Psychology and Department of Psychometrics and Statistics
University of Groningen, The Netherlands

Grant number: FA9550-16-1-0043

Period of performance: November 2017 - November 2018

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Summary

The main progress of 2018 was threefold: (1) another wave of data collection was completed; (2) a conference paper for the annual meeting for the Cognitive Science society was written and submitted (but not accepted for presentation after peer-review); and (3) a paper was written and submitted to a special issue of *Frontiers in Emergency Medicine* (now accepted for publication).

More details regarding those points:

(1) Participants that were tested in previous sessions returned to the lab to assess their CPR proficiency again. Data collection efforts proceeded in parallel with the development of the predictions for performance generated by Predictive Performance Equation.

(2) The conference paper submitted to CogSci2018 detailed the efforts outlined in (1); the experimental paradigm and the computational model were described in detail and the paper discussed the quality of the model predictions. The conference paper was not accepted for presentation at the conference but preparing the conference paper has helped subsequent efforts to document the work in detail and will be used as the basis for a manuscript that we plan to finalize and submit in the next period. This work was done and will be continued in collaboration with Mike Krusmark and Kevin Gluck at the AFRL.

(3) A separate paper was written specifically about the poor performance we observed when participants that have had previous training (months to years before) first entered the lab and how quickly those participants were able to perform proficiently again. This work was submitted to a special issue titled "Keeping Bystanders Active" and co-authored by Kevin Gluck (AFRL). The paper is accepted but not yet published.

Additionally, Van Rijn, Sense, and Maass visited Dayton, OH during the summer to meet face-to-face with Kevin Gluck to discuss the status of the on-going sub-projects. Plans were made to discuss future publications: the first is now complete (see (3)) and two additional papers are planned for the next period. For one of them, the necessary data is already collected and Sense and Maass will spearhead the writing; for the second, an additional wave of data collection was planned for mid-2019.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Predicting Performance in a Complex Real-Life Task: Learning and Forgetting CPR

Sarah Maass* (s.c.maass@rug.nl) & Florian Sense* (f.sense@rug.nl)
Department of Experimental Psychology, University of Groningen
Groningen, The Netherlands
* *Shared first authorship, alphabetical order*

Michael Krusmark (michael.krusmark.ctr@us.af.mil)
L3 Technologies at Air Force Research Laboratory, Wright-Patterson Air Force Base
Dayton, Ohio, USA

Kevin Gluck (kevin.gluck@us.af.mil)
Air Force Research Laboratory, Wright-Patterson Air Force Base
Dayton, OH, USA

Hedderik van Rijn (d.h.van.rijn@rug.nl)
Department of Experimental Psychology, University of Groningen
Groningen, The Netherlands

Abstract

In this study we investigate learning and memory in the real world task domain of cardiopulmonary resuscitation (CPR), a basic lifesaving skill that taps into a combination of procedural and declarative learning. Participants all had previous CPR training months to years earlier, creating an unusual opportunity to assess longer-term retention and model parameterization. In the first experimental session, participants' current CPR skill level was assessed, then they were re-trained and performed several bouts of CPR. The goal was to predict performance levels after 8 and 16 weeks of this training, using the Predictive Performance Equation (PPE; Walsh, Gluck, Gunzelmann, Jastrzembski, & Krusmark, in press). With learning and retention periods spanning months to years, it allowed us to assess the predictive accuracy of PPE over an ecologically relevant timeframe.

Keywords: skill acquisition; retention; learning; memory; CPR

Introduction

Human memory entails both procedural and declarative knowledge. In the lab setting, researchers often isolate either component and study it in controlled conditions. Most real life behaviors and skills, however, draw on complex interactions of memories for both procedural and declarative components. Research detailing the mechanics of procedural or declarative knowledge in isolation might therefore not directly transfer to situations in which both are involved. Hence it is important to study complex real-life tasks in controlled lab settings in order to improve our understanding of the relationship between these types of knowledge.

Cardiopulmonary resuscitation (CPR) is taught in the context of first aid training. It is a useful domain for studying the interaction of procedural and declarative aspects of memory because clear performance standards can be applied to quantifying learners' proficiency. CPR in terms of Basic Life Support, as laid out by the European Resuscitation Council (ERC) guidelines (Soar et al., 2015), includes an assessment of the so-called victim (check response, check breathing) and a series of steps (alert emergency services, hand positioning) before the actual compressions and rescue breaths are administered in a cycle of 30 compressions and two rescue breaths. Therefore, it combines declarative knowledge of which steps to perform and in which order as well as procedural knowledge of how to administer compressions and rescue breaths correctly. Using a modern instrumented Laerdal QCPR manikin, relevant performance measures are recorded and scored.

Crucially, CPR is a skill that needs to be re-trained periodically. For medical professionals, this is typically done every other year - an interval that is often considered suboptimal (Stross, 1983). Furthermore, the American Heart Association (AHA) recognizes that large individual differences exist, which complicates the prescription of ideal methodology and frequency of training (Nolan et al., 2015). Consequently, the acquisition and retention of CPR as a skill provides the ideal testbed for cognitive-psychological theories of learning with the potential of significant practical impact.

The ultimate goal of any psychological theory is to predict behavior. The ability to predict future CPR performance based on behavior recorded at previous training moments is an especially interesting case because it presents the opportunity to predict, for each individual learner, when someone should return to be re-trained.

Collecting data in ecologically valid time frames, however, is extremely time consuming. Using Germans with driver's licenses as research participants can cut the time investment: In Germany, a basic first aid training (including CPR training) is mandatory to obtain a driver's license. German students with driver's licenses enrolled at the university are an interesting sample because we know they have had CPR training in the past and there will be natural variation in how

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

long ago that training was completed. Importantly, there is no mandatory follow-up training and many people will not refresh their knowledge unless required in another context.

Mathematical models of skill acquisition and retention can help describe fluctuations in CPR performance over time. One such model, the Predictive Performance Equation (PPE), formalizes the temporal dynamics in a series of mathematical equations (Walsh, Gluck, Gunzelmann, Jastrzembski, & Krusmark, in press). First, performance increases with the amount of practice (*N*). This is implemented as a power law of learning, with the *learning rate* fixed at 0.1 on the basis of prior empirical evidence and model fits. Second, performance drops as a function of elapsed time among practice events (*T*). This is implemented as a power function of forgetting, with the decay rate determined by the function expressed in Equation 4, below. In PPE the effects of learning and retention on activation is multiplicative, such that:

$$activation = N^{learning\ rate} \cdot T^{-decay\ rate} \quad (1)$$

Third, PPE captures the spacing effect, such that retention is better and more stable when practice is distributed over time. This is implemented in the forgetting function through *T* and the *decay rate*. *T* is computed as the sum of the weighted age of each practice event,

$$T = \sum_{i=1}^{n-1} w_i \cdot t_i, \quad (2)$$

where the weight is an exponential decay function of time,

$$w_i = t_i^{-x} \sum_{j=1}^{n-1} \frac{1}{t_j^{-x}}. \quad (3)$$

Thus, *T* weighs practice repetitions so that more recent events carry more weight, and the variable *x*, which is fixed at 0.6, controls the degree of the weighting.

The *decay rate* is computed as a function of the complete history of lags between successive practice opportunities (lag_j):

$$decay\ rate = decay\ intercept + decay\ slope \cdot average\left(\frac{1}{log(lag_i)}\right) \quad (4)$$

Finally, in PPE performance is computed as a logistic function of activation:

$$Performance = \frac{1}{1 + exp\left(\frac{threshold - activation}{scalar}\right)} \quad (5)$$

PPE parameters are estimated separately for individuals based on their performance histories. For each individual, we compute the optimal values of the *decay intercept*, *decay slope*, *threshold*, and *scalar* that minimizes the mean squared deviation between model and human performance. These parameter values are then used to generate out of sample predictions of performance at future points in time.

An increasingly common modeling practice in sparse data environments is to seed parameter values with priors. There has been some previous exploration of this approach in the context of PPE (Collins, Gluck, Walsh, & Krusmark, 2017). Thus, to enhance the quality of our predictions of CPR performance on Session 2 in the current study, we utilized priors generated from a longer-term CPR study that is currently underway (Jastrzembski, et al, 2017). In initial analyses from this study, CPR compression data from four sessions separated by either one day, one week, one month, or three months were used to calibrate the model and generate model parameters for temporal predictions at either three or six months in the future. These model parameters were utilized in the current study to inform parameters generated during the model fitting process.

The PPE is used to model and predict changes in performance between sessions based on the time that elapsed. Using similar mechanics/equations, one can also optimize repetition schedules *within* individual learning sessions (Sense & Van Rijn, submitted). Earlier work has shown that an ACT-R based cognitive model can use response accuracy and latency on a trial-by-trial basis to predict when each studied item is likely to be forgotten and ensure rehearsal before that moment. This improves retention of the studied facts (Sense & Van Rijn, submitted) and allows the estimation of a learner's rate of forgetting. The adaptive fact-learning model developed in our lab estimates a decay term for each item that each learner studies and by averaging across the item-level parameter values at the end of a study session, a rate of forgetting can be extracted for each participant, indicating how quickly, on average, they forget the studied material. Here, we will test whether the rate with which participants are estimated to forget factual knowledge is related to their decrease in CPR performance measured during the course of the current study.

In addition, we administered the Raven matrices task, which is frequently used as a rough estimate of general cognitive ability (Raven, 2000). If performance on the Raven matrices task is related to someone's ability to acquire and retain CPR knowledge, such a score could provide a potential covariate to refine the predictive model. If not, we would conclude that someone's cognitive ability does not help predict performance in this particular real-life skill.

The current study was devised to assess the accuracy of personalized performance predictions. The ERC's guidelines (Soar et al., 2015) state that "*The intervals for retraining will differ according to the characteristics of the participants*". The availability of predictive models that take an individual's performance profile over time into account permits such personalized predictions. Ideally, this would make re-trainings more efficient and reduce the amount of time that a medical professional would perform suboptimally if necessary. The viability of personalized refresher schedules

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

crucially depends on the accuracy of the predictions: Requiring people to retrain too early is a waste of resources but requiring them too late can cost lives.

Methods

Participants

A total of 50 participants (37 female; 74%) with a median age of 20 (SD = 1.84; range = [18, 27]) were recruited from the participant pool at the University of Groningen. Of those, 10 dropped out after the first session. Requirement for recruitment was the possession of a valid German driver's license. Two participants indicated minimal familiarity with Swahili, which is relevant to the paired associates word learning task in this study. All participants gave informed consent and the Ethics Committee Psychology approved the study (ID: 17017-S-NE(a,b)).

Procedure and Stimuli

Participants completed three experimental sessions. CPR performance was assessed in all sessions and computerized tasks were administered in Session 1 and 2. A graphical summary of the paradigm is provided in Figure 1 and its components are described in more detail below.

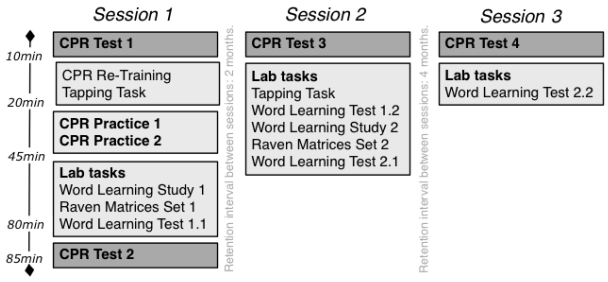


Figure 1. Overview of design and measurement moments.

CPR Test 1 At the beginning of each session, participants signed the informed consent forms. In the first session participants then entered the experimentation room where a Laerdal Resusci Anne QCPR manikin was lying on the ground. Participants read the following instructions: "*You volunteered for community service to help elderly neighbors with chores in their homes. When you enter the house of Mr. Johnson, you find him on the living room floor. There are no signs of bleeding or open wounds and no one else is in the house. Based on your first aid training, take the steps necessary in this situation on the manikin to* ***assess*** *and* ***react upon*** *the situation*."

This scenario was chosen to sketch a hypothetical scenario that required them to perform CPR on the manikin . Specifically, it stated that they are alone, making it necessary to act and not call additional first aid helpers. Any bleeding and wounds were excluded to ensure that participants do not waste time to check for such. Participants were then asked to perform the necessary actions required according to the ERC guidelines (Soar et al. 2015). This means, participants were supposed to alternate between 30 compressions and two rescue breaths. Participants were stopped after administering four cycles of compressions and rescue breaths (i.e., 30-2, 30-2, etc.) to avoid fatigue. We refer to this procedure (i.e., initial steps followed by four rounds of 30-2) as one run-through of CPR.

The initial steps were scored using a pen and paper checklist and the administration of compressions and rescue breaths were recorded on the manikin's SimPad. Performance scores for chest compressions and rescue breaths were based on Laerdal's proprietary scoring algorithm using the European guidelines (ranging from 0 to 100%; a score of 7%5 or higher is considered "proficient"). The SimPad outputs overall scores for compressions, ventilations, and a combined score. The data presented here will focus on the combined scores. This procedure was identical for CPR Tests in all sessions.

CPR Re-Training After the initial assessment, participants were re-trained. First, participants watched a short instructional video (see https://osf.io/9er6g/) demonstrating the initial steps and how to perform them correctly, as well as instructions on how to correctly apply chest compressions and rescue breaths. This video was specifically made for this research project.

Subsequently, participants had the opportunity to practice compressions on the manikin with its live feedback option enabled for one minute. That is, for each compression participants could track their depth and frequency and adjust if necessary. Then, participants practiced giving rescue breaths until the live feedback indicated that two correct breaths had been given in a row.

Tapping Task Participants completed five minutes of the Continuation and Synchronization Tapping Task (Semjen, Schulze, & Vorberg,2000). Results from this task are not relevant in the current context and will not be discussed further.

CPR Practice 1 and 2 Participants were instructed to "*Perform the complete procedure you saw in the video, with four rounds of compressions and rescue breaths*" twice. That is, participants were asked to perform two run-throughs of CPR and their performance was scored as detailed in the previous sub-section.

After the run-throughs of CPR, participants completed questionnaires to gather demographic information, the date their driver's license was issued, and the approximate number of months netween obtaining their license and completing their first aid training. The time between the mandatory training and obtaining the driver's license ranged from 1 to 60 months (mean = 9.92 and SD = 12.71).

Word Learning Study Participants spent 15 minutes studying 35 Swahili-English word pairs using the adaptive method outlined in the Introduction. The word pairs were taken from Nelson and Dunlosky (1994) and items did not repeat across sessions.

Raven Matrices Participants performed a shortened version of Raven's advanced progressive matrices (Raven,

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

2000). Instead of seeing all 36 matrices without a time limit, participants had 10 minutes to solve half of them (as in Foster et al., 2015). Participants saw the 18 odd-numbered items in the first session and the 18 even-numbered items in the second session.

CPR Test 2 Following the computerized tasks, participants were asked to complete another complete run-through of CPR. If the score of this test was below 75%, participants were re-trained until they reached criterion.

Word Learning Test The first session ended with a retention test of the Swahili words the participants learned during Word Learning Study (Test 1.1). Participants were asked to fill in all of the English translations they remembered when presented with the Swahili word. In the second session participants completed the same word retention test (Test 1.2), in addition to the word retention test for the stimuli set of Session 2 (Test 2.1), which will be tested again in Session 3 (Test.2.2).

Predicting Future CPR Performance

All participants will have completed first aid before obtaining their driver's license (in accordance with German law). The approximate date of this initial training is known but no performance measures are available. Hence, all participants' performance was assumed to be at passing rate (i.e., 75%) at that point in time. The scores of CPR Test 1 and 2 in the first session and the time point of the initial training with the corresponding score of 75% were used to make a prediction for the second session using PPE (see Introduction) . These predictions can be compared with the actual performance on CPR Test 3 (see Figure 3). Then the new data point from CPR Test 3 will be added to the model to make predictions for CPR Test 4.

Results

Here we provide descriptive statistics and visualizations of the data through the end of Session 2. Data collection for the third session is scheduled for May/June 2018 and once the complete dataset is available, specific hypotheses and ideas will be tested more formally. In the current context, we will focus on the following: CPR performance at the different test moments, PPE predictions for Session 2 and how they match the collected data, and the correlations between the lab tasks and the CPR components.

The overall CPR performance assessed in the first two sessions is shown in Figure 2. Shown are the scores exported from Laerdal's proprietary software, which combines the compression and rescue breath performance into a single score. The area corresponding to "proficient" performance is shaded green. Measurements from a single participant are grouped by grey lines and color-coded. Overlaid are "violins" to summarize the distribution of scores at each time point.

On the initial test at the beginning of the first session, only two participants demonstrate proficient performance, while many score below 25%. The CPR Re-Training administered between CPR Test 1 and CPR Practice 1 results in a marked increase in performance, making the majority of participants proficient. Testing for a difference between those two scores with a paired Bayesian t-test yields an enormous Bayes factor of 2.4×10^{17} in favor of a difference. The second practice marks a further increase in overall performance and the vast majority of participants retain above-threshold performance until CPR Test 2 at the end of the first session. In the eight week interval until the second session, and performance decreases (BF_{H1} = 9.95) but many participants still exhibit near-ceiling performance.

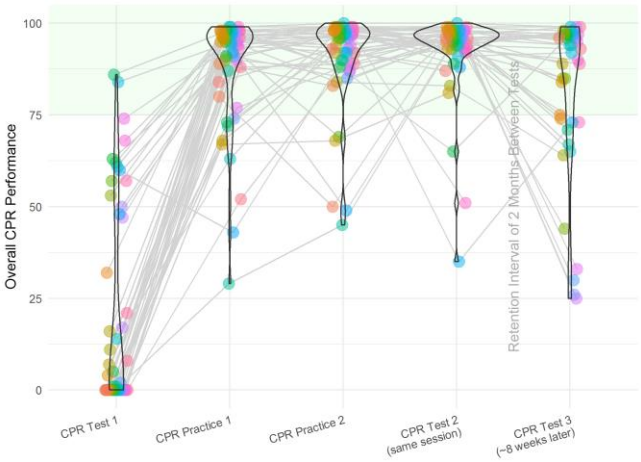


Figure 2. Overall CPR performance at the different measurement moments highlighted in Figure 1.

The recorded performance at the beginning of Session 2 (CPR Test 3 in Figure 2) can be contrasted with the predictions generated using PPE. A graphical summary is provided in Figure 3. The two quadrants shaded in green contain participants for whom predicted and actual performance match relative to the 75% proficiency cut-off. The yellow and red quadrants contain participants for whom performance was under- or over-predicted, respectively. Predictions were made for all 50 participants but actual performance scores are only available for 39 participants in Session 2. Of those, the model predicted that 11 participants would perform above threshold in the second session but a total of 27 did. Conversely, the model predicted that 28 participants would perform below threshold but only 12 did. The predictions and actual performance are highly correlated (r = 0.592) despite the root-mean-square error being relatively high at 32.1. The high error is caused by the cluster of participants in the yellow region, who perform near-perfect while the model predicted scores between 25 and 75%.

The two relevant lab tasks - Raven Matrices and Word Learning - were administered twice each (cf. Figure 1). For the Raven Matrices task, participants saw an average of 16.1 out of 18 problems in the first and an average of 16.4 in the second session. Of those problems, they solved an average of 9.3 in both sessions.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

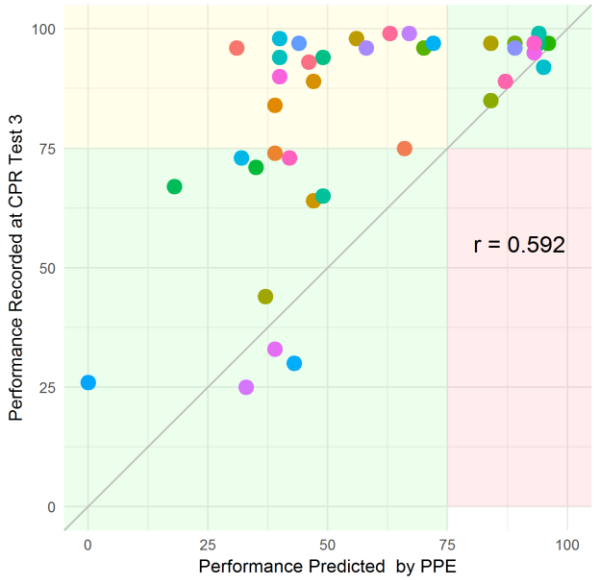


Figure 3. Comparing predicted performance with recorded performance for participants that returned for Session 2.

For the Word Learning task, participants studied an average of 21.3 and 22.2 items (out of 35) in the two sessions, respectively. Each participant's rates of forgetting (one per session) were estimated by averaging across the final value of the estimated parameter associated with each item they encountered at least three times.

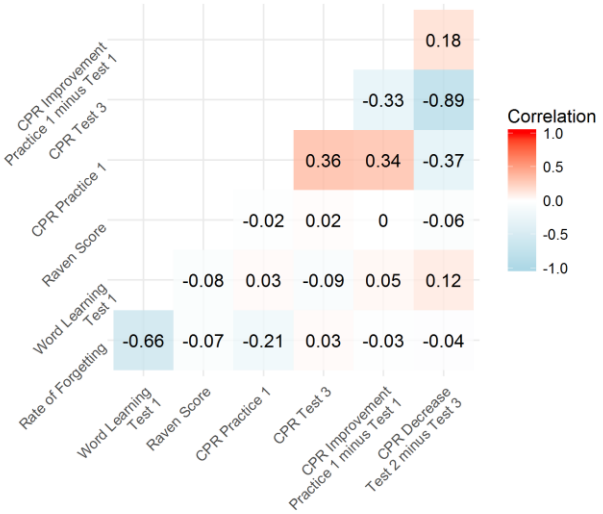


Figure 4. Color-coded correlation matrix for a subset of the relevant outcome measures.

A more detailed overview of the correlations between some of the recorded measures is provided in Figure 4. Given the limited space in the current format, we will focus here on a selection of measures and their correlations. Specifically, we will focus on performance measures from the lab tasks from the first session: the rate of forgetting estimated during Word Learning, the score on the recall test of those same words, as well as the number of Raven matrices solved. These are correlated with four data points that summarize someone's CPR performance: scores on the first practice (i.e., after seeing the instructions), CPR performance after a 2-month delay (i.e., CPR Test 3), as well as the two difference scores corresponding to the Bayes factors reported above. The first captures CPR Improvement from the initial assessment to performance after having received instructions, and the second captures CPR Decrease between the last assessment in first session (CPR Test 2) and the first assessment in the second session (CPR Test 3, see Figure 2).

It is notable that a lot of the correlation coefficients are very small (i.e., < 0.10). Specifically, the Raven scores do not seem to correlate with any of the other measures and neither do the scores on the delayed recall test from the Word Learning task. The estimated rate for forgetting correlates highly with performance on the recall test and the negative correlation indicates that a faster rate of forgetting results in lower test scores (as expected).

The correlation coefficients between the different CPR scores, on the other hand, are slightly more substantial. Most striking is the strong negative relationship (r = -0.89) between someone's performance in Session 2 (CPR Test 3) and the decrease in performance between the first two sessions, which makes sense because someone scoring low on CPR Test 3 must have decreased in performance (also see Figure 2). Similarly, participants whose CPR performance improves tend to score higher on the first practice (r = 0.34) and subsequently also tend to score higher in the second session (CPR Test 3; r = 0.36).

Interestingly, the correlation between the estimated rate of forgetting and the decrease in performance between the first two sessions is almost zero (r = -0.04), suggesting that the rate with which a participants is estimated to forget factual knowledge (within a learning session) is not related to their decrease in CPR performance (between sessions).

Discussion

The main focus of the current work is an attempt to predict future cardiopulmonary resuscitation (CPR) performance using previous CPR performance. Additional lab tasks were administered to explore whether they are related to CPR performance and could be used to further improve performance of future CPR performance. After completion of the first session, predictions have been made for CPR performance in the second session, using performance in Session 1 along with approximated performance and the time between initial training and Session 1. Figure 3 contrasts predicted and recorded performance for Session 2. Now, data on performance for Session 2 are available and can be incorporated in PPE to yield predictions for Session 3.

As can be seen in Figure 3, the predictions are not perfect. In fact, there is a large group of participants that were predicted to perform below the proficiency threshold but performed well above it (i.e., the yellow region). Importantly, however, there are no participants in the red region of the figure: No participant was falsely predicted to be proficient.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

From a model-building perspective, a model making unbiased predictions would be preferred over a biased model. With the focus on a real-life task that is intimately connected to basic lifesaving skills, however, there is an argument to be made for erring on the conservative side. Once we have collected the data for Session 3 and can compare it with the predictions made using the currently available data, it will be very interesting to see whether the accuracy (e.g., RMSE) of the predictions will increase (which we expect) and whether the bias will persist or whether prediction errors will fall in the red region of Figure 3 as well. The prediction is made easier by the availability of more data but will also be made for a retention interval that is twice as long. A practical question that will also remain open until more data are collected is the number of measurement moments that are sufficient to make predictions that are acceptably reliable in real-world contexts. Only then can schedules for personalized refresher trainings be relevant in practice.

The lab tasks were chosen to assess a general measure of cognitive ability (Raven matrices) and a measure of declarative learning (Word Learning). Both tasks are a lot more constrained than the CPR performance measures that capture someone's skill. The lack of correlations between the lab tasks and the various CPR scores (see Figure 4) suggests that performance on the simpler tasks does not predict performance on the more complex tasks. More detailed analyses will be conducted once the complete dataset is available. However, the lack of correlations between the components reported here suggests that future CPR performance is best predicted using previous CPR performance. The very low correlations also suggest that someone's performance on simple lab tasks is not necessarily predictive of their performance on complex real-life tasks.

The ERC's report (Soar et al., 2015) states that CPR performance is known to deteriorate within months of training and, therefore, even annual re-trainings might not be frequent enough for some people. Since CPR training is expensive and optimal training intervals are currently unknown, they suggest that frequent "low dose" training using video instructions and hands-on practice can be as effective as instructor-led courses (Nolan et al., 2015). The work presented here confirms that there is a swift improvement in performance after such CPR Re-Training (cf. CPR Test 1 vs CPR Practice 1 in Figure 2)

In summary, we report an experimental setup in which the learning and forgetting of CPR is assessed over ecologically relevant timeframes. We test a mathematical model's ability to predict future CPR performance using very sparse data. A first wave of predictions is presented here and the second will be presented at the conference. We anticipate predictive accuracy to improve for Session 3 and will be able to evaluate the feasibility of personalized schedules for refresher courses.

Acknowledgments

This work was supported by EOARD grant #11926121.

References

Collins, M. G., Gluck, K. A., Walsh, M. M., & Krusmark, M. A. (2017). Using prior data to inform initial performance predictions on individual students, In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*.

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, 43(2), 226-236.

Jastrzembski, T., Walsh, M. M., Krusmark, M., Kardong-Edgren, S., Oermann, M., Dufour, K., Millwater, T., Gluck, K. A., Gunzelmann, G., Harris, J., & Stefanidis, D. (2017). Personalizing training to acquire and sustain competence through use of a cognitive model. In D. D. Schmorrow & C. M. Fidopiastis, *Augmented cognition. Enhancing cognition and behavior in complex environments* (pp. 148-161). Switzerland: Springer International Publishing AG.

Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory, 2*(3), 325-335.

Nolan, J. P., Hazinski, M. F., Aickin, R., Bhanji, F., Billi, J. E., Callaway, C. W., ... & Gent, L. M. (2015). Part 1: Executive Summary: 2015 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science with Treatment Recommendations. *Resuscitation*, 95, e1-e31.

Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cognitive Psychology*, 41(1), 1-48.

Semjen, A., Schulze, H. H., & Vorberg, D. (2000). Timing precision in continuation and synchronization tapping. *Psychological Research*, 63(2), 137-147.

Sense, F., & Van Rijn, H. (submitted). Optimizing Within-Session Repetition Schedules with a Response-Latency-Based Cognitive Model Improves Retention.

Soar, J., Nolan, J. P., Böttiger, B. W., Perkins, G. D., Lott, C., Carli, P., ... & Sunde, K. (2015). European Resuscitation Council guidelines for resuscitation 2015. *Resuscitation*, 95, 100-147.

Stross, J.K. (1983). Maintaining Competency in Advanced Cardiac Life Support Skills. *Journal of the American Medical Association*, 249(24), 3339–3341.

Walsh, M. M., Gluck, K. A., Gunzelmann, G., Jastrzembski, T., & Krusmark, M. (in press). Evaluating the theoretical adequacy and applied potential of computational models of the spacing effect. *Cognitive Science*.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

frontiers

Keeping Bystanders Active: Resuscitating Resuscitation Skills

Sarah C. Maaß*[1,2,] **, Florian Sense*[1,2], Kevin Gluck[3], & Hedderik van Rijn[1]

*** *shared first authorship***

[1]Department of Experimental Psychology, University of Groningen, The Netherlands
[2]Behavioral and Cognitive Neurosciences, University of Groningen, The Netherlands
[3]Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, USA

**** Correspondence:**
Sarah C. Maaß
s.c.maass@rug.nl
Number of words: 2640
Number of figures: 2
Number of tables: 1

Keywords: Cardiopulmonary resuscitation$_1$, active bystanders$_2$, long-term retention$_3$, low-intensity retraining$_4$, general population proficiency$_5$, video and simulation training$_6$.

Abstract

Objective. Sufficient CPR skills in the general population are essential to make them active bystanders and contribute to an effective rescue chain in cardiac arrest emergencies. However, having a large proportion of the population be regularly retrained is practically infeasible. The aim of this study was to assess and retrain cardiopulmonary resuscitation (CPR) skills of individuals who received (limited) CPR training several months to years prior.
Method. 99 German adults in a possession of a driver's license were asked to perform CPR on a Laerdal Resusci Anne® QCPR manikin (Laerdal, Stavanger, Norway). After initial assessment, participants watched an instructional video and completed short, isolated compression and ventilation practice with live feedback. CPR proficiency was assessed again after retraining and after a retention interval of 45 minutes.
Results. Our results indicate that only 2% of participants managed to reach the performance criterion set by the European Resuscitation Council Guidelines, with most failing to reach even the lowest levels of performance. This corroborates earlier observations that CPR skills have deteriorated almost completely after a long retention interval, calling into question "one-and-done" certification of this basic life-saving. However, we also demonstrate that performance strikingly increased after watching a 6-minute instructional video and a short opportunity for isolated practice. This increase in performance was stable over 45 minutes with 96% of participants meeting criterion performance Closer inspection of the isolated compression practice data suggests that performance was very high at the start of the practice already, indicating that short refresher videos might suffice to change bystanders that would not have initiated CPR due to lack of knowledge into active first responders.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Conclusion. We suggest that short refresher trainings could be an effective and affordable means of improving basic lifesaving skills to increase the effective contribution of bystanders during emergencies.

1 Introduction

With 70,000 to 80,000 individuals per year experiencing sudden cardiac arrest in Germany, even without major, wide-scale emergencies, this condition is among the country's most frequent causes of death. As the time between cardiac arrest and starting cardiopulmonary resuscitation (CPR) is critical for survival, effective bystander CPR is a key determinant of the survival rate of cardiac arrest victims [1]. Sufficient CPR skills in the general population are essential for an effective rescue chain [2]. Even though CPR-trained bystanders might still fail to perform CPR [3], lacking even basic CPR knowledge makes it impossible for any bystander to actively contribute in cases of emergencies: Even though bystanders might not meet the proficiency levels of skilled medical personnel, attempting to start CPR is better than inactivity among bystanders [4].

In Germany, a large percentage of the adult population has been trained in resuscitation (over 80% in a sample of 200 participants [5]) because completing CPR training is part of the requirements for obtaining a driver's licence. As this is a one-session First Aid certification, including CPR training, no retraining or recertification is formally required. However, given the population level proficiency, it is surprising that an alarmingly low percentage of people provide First Aid care, and more specifically, initiate CPR before Emergency Medical Services arrive [5]. Given the critical role of bystanders in minimizing the consequences of a cardiac arrest, a role that would be even more critical in the case of large-scale emergencies, this discrepancy between a potentially well-trained population and low levels of CPR assistance is worth investigating.

As CPR skills are known to deteriorate over time [6], it could be that the CPR training that is part of the one session First Aid certification is not sufficient for long-term proficiency. This idea is supported by the observation that even in medical professionals performance typically deteriorates in the first year after certification [7,8]. Furthermore, in a recent German study, only 40% of a general-public sample reported they were capable of initiating CPR, as measured via a self-assessment questionnaire [5]. This percentage increased markedly after a short refresher training was provided. No quantitative performance data were reported in that study, but it suggests that additional instruction at least increases subjectively perceived CPR skills.

To quantify the proficiency of CPR skills in the general population and the potential effectiveness of a short, low-intensity refresher training to reduce any knowledge-based constraints in performing CPR, we first assessed the level of CPR performance that was retained over months to years after a first-aid course. Second, we assessed the increase in performance after a short training video used to restore CPR proficiency to European Resuscitation Council Guideline standards [9].

2 Methods

For this study, we invited 99 German participants (age 18-27, M = 20.8, enrolled as students at the University of Groningen, The Netherlands) with valid German driver's licenses. The time since obtaining their driver's license ranged from less than a month to more than 10 years prior to participation (M = 45.8, SD = 26.05, in months). All participants gave written informed consent and ethical approval was obtained to conduct the study (WIRB protocol #20151567).

This is a provisional file, not the final typeset article

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Immediately after arriving for the study, participants underwent an initial skill assessment in which they performed CPR on a Laerdal Resusci Anne® QCPR manikin (Laerdal, Stavanger, Norway). Performance was recorded using a SimPad PLUS device (Laerdal, Stavanger, Norway) with activated SkillReporter software, without any form of feedback. After this initial assessment, all participants watched a 5:46 minute instructional video (see https://osf.io/w5ze2/) in which the basic CPR procedure, following European Resuscitation Council Guidelines 2015 [8], was explained and demonstrated. Immediately after this video, they were asked to perform compressions and rescue breaths independently for one minute each, with the manikin's live-feedback option enabled on the SimPad and thus providing immediate feedback to the participant. After these training sessions, CPR performance was re-assessed in two tests. During these tests, no feedback was provided, but a certified CPR trainer provided personalized feedback and instruction in between these tests when necessary. After a delay of 45 minutes, in which participants performed a number of unrelated tasks, CPR skills were assessed again using the SimPad device, without feedback. Note that this study was embedded in a multi-session experiment, details of which will be reported in a separate manuscript (see https://osf.io/zpjrt/ for project details).

3 Results

On the initial skill assessment, 78.4% of participants did not attempt to administer rescue breaths and 6.2% did not perform a single compression (i.e., did not initiate CPR at all). Even though most participants did initiate chest compressions, the vast majority ceased chest compressions too early: The average number of compressions was 48.2, with 23.7% administering less than 20 and 83.5% less than 90 compressions. This results in the extremely poor overall CPR performance on the initial skill assessment evident for many participants in Figure 1: 50.5% of participants score 0 and 75.3% score below 20. That is, participants' initial performance was far below minimum proficiency levels, with only 2% of participants meeting the criteria set by the European Resuscitation Council Guidelines of 2015 [8] (first column in Figure 1).

Figure 1: Individual participants' (n=99) composite CPR performance scores and violin plots of the sample distribution, measured with the Laerdal Resusci Anne® QCPR manikin (Laerdal, Stavanger, Norway) during four test sessions. Green area indicates performance above the minimum proficiency level [8]. Numbers above the violin plots indicate percentage of participants performing this skill above the minimum required proficiency level of 75.

Interestingly, skill performance improved strikingly after watching the instructional video and performing compressions and rescue breaths for one minute, with 81% of participants passing the criteria set by the European Resuscitation Council after this short refresher (second column in Figure 1). Personalized feedback further improved performance (third column, Figure 1), a performance level which seemed to be stable when retested after 45 minutes (fourth column, Figure 1). Importantly, participants also markedly improved with respect to the required steps that have to be performed before executing CPR, which are listed in the four top-most rows of Table 1.

As suggested previously by Van Berkom et al. [10], there may be some cause for concern regarding the composite performance score generated by the embedded measurement system. However, the measurement system on the manikin also records the separate measures comprising the composite score - such as compression rate, compression depth, release height, and hand position -, which can be analysed instead of the proprietary composite score. Analyses of these separate measures indicate that performance at the start of the training sessions was already at the performance level required by

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

the European Resuscitation Council Guidelines, indicating that watching the short video, without hands-on training, improved CPR performance. This can be seen in Figure 2 in which performance over time during the compression practice training is plotted for each participant separately (each line in the plot represents the values recorded for a single participant). This figure demonstrates that only a handful of participants completely fail to reach the required compression frequency, and that most achieve the 104 compressions per minute demonstrated in the video. Moreover, even though a proportion of the participants require the live-feedback to increase their compression rate to ERC criteria [9], most participants immediately start out with reasonable compression rates. We refer the interested reader to the supplementary analyses (https://osf.io/zpjrt/), which contains equivalent visualizations of other subscores recorded by the manikin. In summary, watching a short instructional video sufficed to retrain either inactive or ineffective bystanders.

Figure 2. Compression frequency (in compressions/minute) per participant as recorded during the one-minute training session immediately after watching the instructional video. Green area indicates performance in line with set criteria [9]. Blue line indicates average performance at each point in time.

4 Discussion

The results of this study demonstrate that: (1) After a long retention interval, even young, highly educated individuals are not "ready, willing and able to act" [11] when confronted with a simulated victim of a cardiac arrest, but that (2) Watching a short instructional video, potentially combined with short practice sessions including feedback, restores performance for the majority of participants. Combined, this work suggests that providing one-shot training to young adults is not sufficient for them to retain basic lifesaving skills which could be required in cases of an emergency. However, merely presenting an instructional video that refreshes previously trained skills may be sufficient to regain proficiency in these basic lifesaving skills for the majority of participants, even years after initial training. These findings are especially relevant if there already is an infrastructure to provide initial training to the general population (like in Germany), and were dormant skills could be efficiently retrained to reach widespread CPR proficiency.

These results also highlight the limited effectiveness of a single training session, when not followed up by later refresher trainings. "One-and-done" certification practices are antiquated and unresponsive to contemporary basic and applied cognitive science [12]. The data presented here also suggest that the proportion of potential bystanders that actually initiate and perform effective CPR (as low as 2%, see Figure 1) might be much lower than the number of people that self-report that they would initiate CPR (40% according to [5]). Our results are consistent with some previously published research [13] suggesting that refresher training can be of much lower intensity, as people with previous training relearn more effectively [14]. The effectiveness of watching an instructional video, as reported here, strongly argues for making brief refresher training easily accessible. That video training can be as effective as instructor-led training has been demonstrated in earlier work [15], in which a 30-minute video-based self-training provided positive results. Interestingly, our work demonstrates that a much shorter, 6-minute video clip can also result in proficient CPR performance.

Providing access to short but effective (re)training will probably lower the threshold for self-initiation of training, or would make it more likely that people would follow-up on externally provided reminders for self-training [16]. Given the almost perfect coverage of CPR training in German adults, one could argue that such instructional videos should be made available via screens in public spaces, or be part of national broadcasting services. If these videos were attended regularly,

This is a provisional file, not the final typeset article

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

the resuscitation skills of an entire population could be continuously resuscitated, increasing the chances that bystanders will act effectively in case of a cardiac arrest in emergencies, potentially saving thousands of lives in the process.

Table 1.
CPR Skill Assessment Scores.[1]

	Initial Skill Assessment	Assessment after Video and 1 min Training Sessions	Second Assessment after Video and Personalized Instruction	Test after 45 min
Check response	55%	100%	100%	99%
Call for help	62%	93%	97%	96%
Open Airway	5%	82%	94%	95%
Breathing check	71%	92%	97%	98%
Proper hand position	91%	99%	100%	99%
Compression	11%	85%	91%	96%
Ventilation	7%	89%	92%	94%

[1] Percentages indicate proportion of subjects performing satisfactorily (criteria described below). For *Check responsiveness*, *Call for help*, *Open airway*, and *Breathing check* the scoring was binary (correct/incorrect scored by a certified CPR instructor) and the data in the table show the % of participants who did the action correctly. For compression and ventilation the percentages indicate the proportion of subjects reaching the criterion of 75% in skill performance recorded with the SimPad PLUS device (Laerdal, Stavanger, Norway). Note that *Proper hand position*, *Compression* and *Ventilation* are part of the composite score reported in Figure 1.

5 References

1. Ritter G, Wolfe RA, Goldstein S, Landis JR, Vasu CM, Acheson A, et al. The effect of bystander CPR on survival of out-of-hospital cardiac arrest victims. *American heart journal* 1985; **110**:932-937.
2. Sasson C, Rogers MA, Dahl J, Kellermann AL. Predictors of survival from out-of-hospital cardiac arrest: a systematic review and meta-analysis. *Circulation: Cardiovascular Quality and Outcomes* 2010; **3**:63-81.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

3. Swor R, Khan I, Domeier R, Honeycutt L, Chu K, Compton S. CPR training and CPR performance: do CPR-trained bystanders perform CPR?. *Academic Emergency Medicine* 2006; **13**: 596-601.
4. Fridman M, Barnes V, Whyman A, Currell A, Bernard S, Walker T, et al. A model of survival following pre-hospital cardiac arrest based on the Victorian Ambulance Cardiac Arrest Register. *Resuscitation* 2007; **75**:311-322.
5. Malsy M, Leberle R, Graf B. Germans learn how to save lives: a nationwide CPR education initiative. *International journal of emergency medicine* 2018 **11:**9.
6. Soar J, Monsieurs KG, Ballance JH, Barelli A, Biarent D, Greif R, et al. European resuscitation council guidelines for resuscitation 2010 section 9. Principles of education in resuscitation. *Resuscitation* 2010; **81**:1434-1444.
7. Mancini ME, Kaye W. The effect of time since training on house officers' retention of cardiopulmonary resuscitation skills. *The American journal of emergency medicine* 1985; **3**:31-32.
8. Smith KK, Gilcreast D, Pierce K. Evaluation of staff's retention of ACLS and BLS skills. *Resuscitation* 2008; **78**:59-65.
9. Perkins GD, Handley AJ, Koster RW, Castrén M, Smyth MA, Olasveengen T, et al. European Resuscitation Council Guidelines for Resuscitation 2015: Section 2. Adult basic life support and automated external defibrillation. *Resuscitation* 2015; 81-99.
10. Van Berkom P, Cloin J, Van Beijsterveldt T, Venama A. The automated Laerdal Resusci Anne scoring system: Scoring on slippery curves, *Resuscitation* 2017; **118**:e96.
11. Nolan JP, Soar J, Zideman DA, Biarent D, Bossaert LL, Deakin C, et al. European resuscitation council guidelines for resuscitation 2010 section 1. Executive summary.*Resuscitation* 2010; **81**:1219-1276.
12. Gluck KA, Jastrzembski TS, Krusmark MA. (in press). Prospective comments on performance prediction for aviation psychology. In M. Vidulich & P. Tsang (Eds.). *Advances in Aviation Psychology*. Cambridge, MA: MIT Press.
13. Sutton RM. Niles D, Meaney PA, Aplenc R, French B, Abella BS, et al. "Booster" training: evaluation of instructor-led bedside cardiopulmonary resuscitation skill training and automated corrective feedback to improve cardiopulmonary resuscitation compliance of Pediatric Basic Life Support providers during simulated cardiac arrest. *Pediatric critical care medicine: a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies* 2011; **12**:e116.
14. Wiese CHR, Bartels U, Bahr J, Graf BM. Durchführung von CPR-Basismaßnahmen durch Laienhelfer nach dem Kurs „Lebensrettende Sofortmaßnahmen" Practical examination of first aid bystanders performing Basic Life Support. *Notfall+ Rettungsmedizin* 2007; **10**: 221-228.
15. Einspruch EL, Lynch B, Aufderheide TP, Nichol G, Becker L. Retention of CPR skills learned in a traditional AHA Heartsaver course versus 30-min video self-training: a controlled randomized study. *Resuscitation* 2007; **74**:476-486.
16. Ahn JY, Cho GC, Shon YD, Park SM, Kang KH. Effect of a reminder video using a mobile phone on the retention of CPR and AED skills in lay responders. *Resuscitation* 2011; **82**:1543-1547.

This is a provisional file, not the final typeset article

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

6 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7 Author Contributions

SCM, FS, KG, and HVR contributed to the conception and design of the study, the planning of the analyses, and the interpretation of the results; SCM performed the data collection. FS organized the database and performed the statistical analyses. SCM and FS wrote the first draft of the manuscript; KG and HVR wrote sections of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

8 Funding

This work was supported by EOARD grant #11926121.

9 Acknowledgments

We would like to thank Michael Krusmark for valuable contributions to the conceptions of the project. We also thank Luzia Heu for her help filming the instructional video

1 Data Availability Statement

The datasets analyzed for this study can be found in the Open Science Framework (https://osf.io/zpjrt/).

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

journal of cognition

Sense, F., et al. 2019 Within-Subject Performance on a Real-Life, Complex Task and Traditional Lab Experiments: Measures of Word Learning, Raven Matrices, Tapping, and CPR. *Journal of Cognition,* 2(1): 12, pp. 1–10. DOI: https://doi.org/10.5334/joc.65

DATA REPORT

Within-Subject Performance on a Real-Life, Complex Task and Traditional Lab Experiments: Measures of Word Learning, Raven Matrices, Tapping, and CPR

Florian Sense[1,2], Sarah Maaß[1,2], Kevin Gluck[3] and Hedderik van Rijn[1,2]

[1] Department of Experimental Psychology, University of Groningen, NL

[2] Behavioral and Cognitive Neurosciences, University of Groningen, NL

[3] Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, US

Corresponding author: Florian Sense (f.sense@rug.nl)

In this data report, we describe a three-session experiment spanning six months. Several well-controlled laboratory tasks (Word Learning, Raven Matrices, and Tapping) and Cardiopulmonary Resuscitation (CPR), a complex but well-defined real-world task, were administered. Data are reported from 50 participants for the first session, 40 for the second, and 34 for the third. CPR is a useful domain for studying real-world performance inside the laboratory because clear performance standards can be applied to quantifying learners' proficiency covering both the first steps that need to be taken prior to the initiation of CPR (declarative knowledge) as well as the compressions and ventilations themselves (procedural skill). This research resulted in a rich dataset with a range of different measures for all participants. For all tasks, the complete set of raw data are made available along with relevant aggregate performance scores (see https://osf.io/m8bxe/). The raw data in particular will enable other researchers to explore potential analyses and modeling beyond the scope of our own. The details of the data collection protocol and available data are documented here to facilitate this process.

Keywords: Long-term memory; Executive functions; Learning; Memory; CPR

Introduction

Psychologists routinely study learning, forgetting, problem solving, and other cognitive tasks of various complexities in the lab. Usually, laboratory tasks are designed to create a tightly controlled environment to scrutinize a single aspect of the process under investigation. We had the opportunity to collect data that combined three such well-controlled, single-purpose tasks (Word Learning, Raven Matrices, and Tapping) along with a complex, real-world task: Cardiopulmonary Resuscitation (CPR). CPR is a useful domain for studying a real-world task inside the laboratory because behavior is highly scripted, with individuals clearly differing in their performance and the speed by which it deteriorates (McKenna & Glendon, 1985), but clear performance standards that can be applied to quantifying learners' proficiency (Perkins et al., 2015). The guidelines describe the steps that need to be taken prior to the initiation of CPR (relying on declarative knowledge), as well as rules for the frequency and force with which compressions and ventilations need to be performed (relying on procedural skills). Due to modern, instrumented CPR manikins, precise data on procedural performance are available, which allows modelling the behavioral profiles at high temporal resolution. The accuracy of the initial steps can be scored by the experimenter so the entire spectrum of CPR performance is quantified.

This research resulted in a rich dataset with a range of different measures for all participants. For all tasks, the complete set of raw data are made available along with relevant aggregate performance scores (see https://osf.io/m8bxe/). In the following, we will provide a brief overview of the administered tasks and

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

what motivated their inclusion. The following sections in this document provide the *data collection protocol* and a complete overview of the *available data*.

We selected three lab tasks that might capture individual differences in the declarative and procedural aspects of CPR performance. The word learning task uses an adaptive learning algorithm that has been shown to be a reliable measure of retention of declarative knowledge (Sense, Behrens, Meijer, & van Rijn, 2016). In addition to the individual responses to each item – when are items presented, accuracy and latency of each response – we also report an overall rate of forgetting per participant, providing an index to their declarative knowledge skills. Earlier work suggests that the rate of forgetting is not simply a function of general cognitive ability or working memory capacity (Sense, Meijer, & van Rijn, 2018), which is why we included a dedicated proxy for general cognitive ability: a variation of the Raven matrices task (Raven, 2000). The synchronization and continuation tapping task (Semjen, Schulze, & Vorberg, 2000) was included to measure participants' capacity to coordinate perception of and action to a rhythmic pattern. Coordinating perception and action is relevant in the context of CPR as compressions need to be administered at a specific and stable rate for them to be clinically effective. We fixed tapping rates at 110 bpm in one block because it is the mean of the recommended compression rate (100–120 bpm). Furthermore, the instructional video showcased compressions being administered to the beat of "Staying Alive" by the Bee Gees (103 beats per minute; bpm) to provide a cue that participants can "synchronize" with when initiating compressions and later need to "continue" during CPR.

In this manuscript, we document in detail how all data were collected. Our goal is to enable other researchers to exploit these data and we believe that they are especially interesting to those interested in within-subject fluctuations in performance, both within and across sessions. The sample size is relatively small, which makes computing correlations between aggregate measures problematic. However, there is a wealth of data for each participant, allowing a range of exploratory analyses. For example, one could assess the external validity of the synchronization and continuation task by linking detailed measures of moment-to-moment variability to the variability observed in the compression data, or determine motor-noise patterns at an individual level to correlate them with the confidence by which memory statistics were determined on the basis of keystroke latency data. In addition, the detailed CPR data (see **Table 1**) lend themselves to an investigation of which types of errors participants make when re-learning CPR and which aspects of performance suffer more than others after a delay: Are participants as likely to forget the correct compression depth as they are to forget the correct hand placement? And are these patterns mediated by intelligence and/or the rate of forgetting? Many other potential questions could be explored and we hope the details in this report facilitate this process.

Data Collection Protocol

In the following we detail the data collection protocol and provide an overview of the data available for each administered task. **Figure 1** gives a visual overview of the administered tasks and the logic of the labelling used.

Table 1: Information available for each recording session of CPR testing and practice. Terms in *italics* correspond to variable labels in the raw data files.

Category (*type*)	Rows	Description of category information	Example *type2* information
CprSessionInfo	14	Info about manikin settings and start and end times of each recording.	*totalTime*
CprSessionStatistics	76	Means and SDs for performance metrics and information for visuals displayed on SimPad.	*compCount, ventMeanVolume, ventTooMuchVolumePercentage*
CprSessionScore	23	Overall score, compression, and ventilation scores, and score reductions applied.	*overallScore, ventOverallScore, compRateHighScoreReduction*
CprInactivityCPR	variable	Coding of the inactivity between compression and ventilation events.	*compInactivity*
CprCompEvent	variable	Timestamped details on each recorded compression.	*compDepth, compReleaseDepth*
CprVentEvent	variable	Timestamped details on each recorded ventilation.	*ventVolume*

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

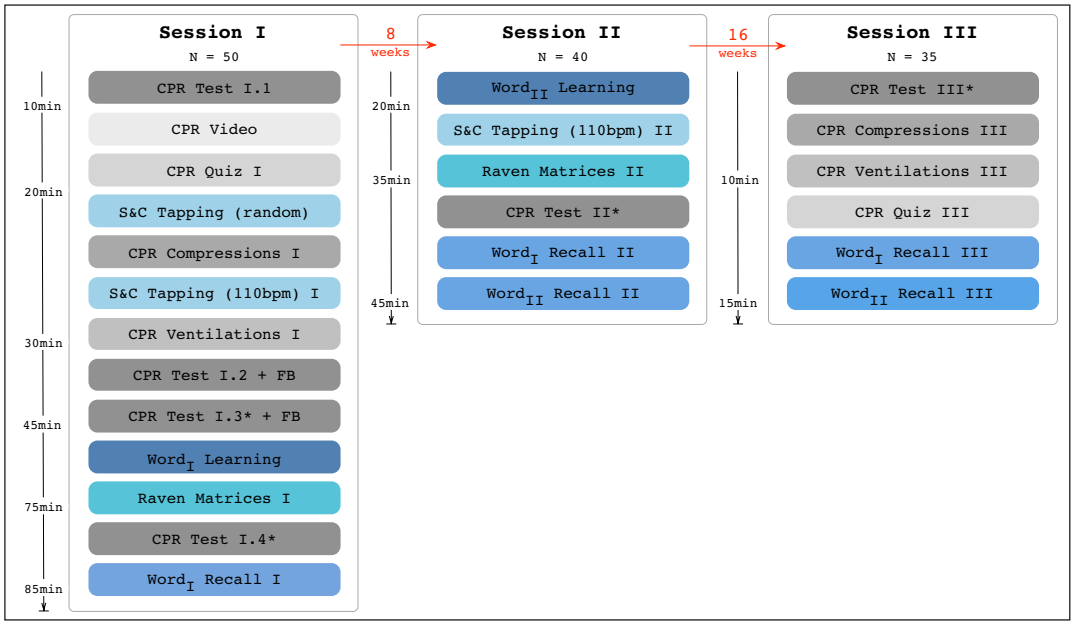


Figure 1: Overview of the data collection protocol. Asterisks (*) indicate that participants were trained until CPR performance was above criterion (>75%).

Participants

A total of 50 participants (37 female; 74%) with a median age of 20 (SD = 1.84)[1] were recruited from the participant pool at the University of Groningen. Of these, 10 dropped out after the first session, and a further 5 after the second, leaving 35 participants with data for all three sessions.[2]

Requirement for recruitment was the possession of a valid German driver's license (because of the requirement in Germany to complete first aid training before obtaining the license—ensuring that participants had been trained previously). Two participants indicated minimal familiarity with Swahili, which is relevant to the paired associates word learning task in this study. All participants gave written informed consent for each session and the study was approved by the Western Institutional Review Board (protocol #20151567) and the local Ethics Committee Psychology (ID: 17017-S-NE(a,b)).

Procedure and Stimuli

Participants completed three experimental sessions with delays of 8 weeks between Sessions I and II, and 16 weeks between Sessions II and III. At the beginning of each session, each participant signed an informed consent form. A graphical summary of the paradigm is provided in **Figure 1**, indicating the timing both within and between sessions. Each of the components are described in more detail below.

Session I

CPR Test I.1

In the first session participants entered the experimentation room where a Laerdal Resusci Anne® QCPR manikin[3] was lying on the ground. Participants read the following instructions: "*You volunteered for community service to help elderly neighbors with chores in their homes. When you enter the house of Mr. Johnson, you find him on the living room floor. There are no signs of bleeding or open wounds and no one else is in the house. Based on your first aid training, take the steps necessary in this situation on the manikin to* **assess** *and* **react upon** *the situation.*"

This scenario was chosen to sketch a hypothetical scenario that required participants to perform CPR on the manikin. Specifically, it stated that they are alone, making it necessary to act and not rely on additional first aid helpers. Any bleeding and wounds were excluded to ensure that participants do not waste time to

[1] Note that in the publicly available data, age has been categorized as <20, 20–24, and 25+ in line with privacy regulations.
[2] Note that the lab tasks were administered in Sessions I and II.
[3] https://www.laerdal.com/sg/products/simulation-training/resuscitation-training/resusci-anne-qcpr/.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

check for these. Participants were tested on their performance of the necessary actions required according to the European Resuscitation Council (ERC) guidelines (Soar et al., 2015). For perfect performance, participants had to correctly remember and complete a sequence of actions: check responsiveness, check for breathing and open airways, alert emergency services, and determine correct hand position for compressions. They then alternated between 30 compressions and two rescue breaths. Participants were stopped after administering four cycles of compressions and rescue breaths to avoid fatigue. We refer to this procedure (i.e., initial steps followed by four rounds of 30-2) as one run-through of CPR.

The declarative steps were scored by the experimenter using a pen and paper checklist. The administration of compressions and rescue breaths was recorded on the manikin's SimPad, a digital tablet device linked with the manikin's sensors. This procedure — scoring of steps taken followed by one run-through of CPR — was identical for all components marked "CPR Test" in **Figure 1** and will be referred to as a *complete* run-through.

CPR Video and Quiz I

After the initial assessment, participants were re-trained. First, participants watched a short instructional video specifically made for this research project (see https://osf.io/9er6g/) demonstrating the initial steps, as well as instructions on how to correctly apply chest compressions and rescue breaths. To test whether participants understood and remembered the CPR procedure they completed a multiple-choice quiz about the material presented in the video.[4]

S&C Tapping (random)

Subsequently, participants completed six trials of the Synchronization and Continuation (S&C) Tapping Task (based on Semjen et al., 2000). One trial consisted of a synchronization phase and a continuation phase. In the first, 32 recurring short beeps (10ms long) were played with an inter-stimulus interval (ISI) randomly sampled per trail (i.e., random on each trial but fixed for the 32 beeps within a trial) from a uniform distribution between 400 and 800ms. The participants had to tap along with the beep on a trackpad until 32 key presses were recorded ("synchronization" phase). Then the beeps stopped and the participants were instructed to continue tapping another 32 times at the same rate ("continuation" phase). Performance was expressed as the difference between the participant's interval between the current and previous tap, and the ISI valid for this trial.

CPR Compressions I

Following this, participants had the opportunity to practice compressions on the manikin with its live feedback option enabled for one minute. That is, while performing chest compressions, participants could track their depth and frequency on the SimPad and adjust if necessary.

S&C Tapping (110bpm) I

Following this, another set of six trials of the S&C tapping task were administered with a fixed ISI of 600ms, which is equivalent to the correct compression rate during CPR (110bpm).

CPR Test I.2 and I.3 + FB

Participants were instructed to "*Perform the complete procedure you saw in the video, with four rounds of compressions and rescue breaths*" twice, while their performance was scored as outlined under *CPR Test I.1*. In addition, participants received verbal feedback after each test. If the score of *CPR Test I.3* was below 75%, an additional run-through of CPR was completed.

After the run-throughs of CPR, participants completed questionnaires to gather demographic information, the date their driver's license was issued, and the approximate number of months between completing their CPR training and obtaining their license. The time between the mandatory training and obtaining the driver's license ranged from 1 to 60 months (mean = 9.92 and SD = 12.71).

[4] Responses on the multiple-choice quizzes are not discussed in more detail here but the data are available in the OSF repository (https://osf.io/m8bxe/).

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

$Word_I$ Learning

Participants spent 15 minutes studying 35 Swahili-English word pairs. The word pairs were taken from Nelson and Dunlosky (1994) and two unique 35-item sets were created, one each for Session I and II. During word learning, word-pairs were introduced in random order and subsequent repetitions were governed by an adaptive scheduling algorithm (van Rijn, van Maanen, & van Woudenberg, 2009) based on a computational model of human memory that – given previous exposure to a fact – estimates the probability of a correct retrieval during word learning (also see Pavlik & Anderson, 2008). Word-pairs were repeated when the algorithm estimated that answers were almost forgotten, thus balancing the benefits of the spacing and testing effects (for technical details, see Sense et al., 2016). The model's parameters are continuously adjusted as responses during word learning are collected and compared to the model's predictions. These item-level model parameters can subsequently be aggregated for each participant and serve as an individual differences measure that we call *rate of forgetting* (Sense et al., 2018). Because the repetitions were adaptively scheduled based on participants' responses during word learning, some participants will not have seen all word-pairs and the number of repetitions per word-pair varies between participants.

Raven Matrices I

Participants performed a shortened version of Raven's advanced progressive matrices (Raven, 2000). Instead of seeing all 36 matrices without a time limit, participants had 10 minutes to solve half of them (as in Foster et al., 2015). Participants saw the 18 odd-numbered items in the first session (*Raven Matrices I*). Materials were taken from (Loesche, Wiley, & Hasselhorn, 2015).

CPR Test I.4

Following the computerized tasks, participants were asked to do another complete run-through of CPR. If the score was below 75%, participants were re-trained (by giving feedback of what needed to improve) and re-tested until they reached criterion (with maximum of two additional iterations, and participants self-paced the start of the re-tests to take breaks if necessary).

$Word_I$ Recall I

The first session ended with a retention test of the Swahili words. The cued recall test listed all Swahili words that could be studied during *$Word_I$ Learning* and participants could respond in any order without a time limit. No feedback was provided.

Session II

Participants spent 15 minutes studying a new set of 35 Swahili-English word pairs using the adaptive method described above (*$Word_{II}$ Learning*). Then, participants completed six trials of the S&C tapping task with a fixed ISI of 600ms (*S&C Tapping (110bpm) II*) followed by the 18 even-numbered items of the shortened version of Raven's advanced progressive matrices (*Raven Matrices II*). Next, participants were asked to do another complete run-through of CPR (*CPR Test II*) which was repeated if performance was below 75%. Lastly, participants completed the same kind of cued recall test for both the words from Session I (*$Word_I$ Recall II*) and Session II (*$Word_{II}$ Recall II*).

Session III

The third session started with another complete run-through of CPR (*CPR Test III*), followed by one minute of compressions (*CPR Compressions III*) without live feedback. Then participants were asked to administer rescue breaths until two consecutive ventilations were correctly performed (*CPR Ventilations III*). Next, the CPR multiple choice quiz from Session I was repeated (*CPR Quiz III*). Finally, participants completed the word retention test for the stimulus set of Session I (*$Word_I$ Recall II*) and the stimulus set of Session II (*$Word_{II}$ Recall II*).

Available Data

Here, we will present in more detail the data available for each task, especially for the CPR components, as these are the most complex and novel. Each of the following sections corresponds to a file in the "Individual Tasks" folder in the OSF repository (https://osf.io/m8bxe/) and the labels used to refer to the components are based on **Figure 1**. The supplement contains more information on the raw data and demonstrates how each component discussed below was computed.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

CPR Performance

All interactions of all participants with the manikin were recorded on the SimPad, resulting in detailed information. Performance scores for chest compressions and rescue breaths were based on Laerdal's proprietary scoring algorithm (ranging from 0 to 100%; a score of 75% or higher is considered "proficient"). We selected "Guideline type Adult" of the "2010 (EU) – ERC" European Resuscitation Council guidelines in the configuration settings of the SimPad, to ensure that the scoring algorithm was based on criteria set by the recommendations of the ERC. The SimPad outputs overall scores for compressions, ventilations, and a combined score. Those scores as well as a plethora of low-level information is stored in six categories, the first three of which are always the same and the last three vary with the number of recorded events. **Table 1** provides an overview and some examples. All participants had previous first aid training. The approximate date of that training and whether they had additional training since was also recorded.

This allows extraction of: High-level performance metrics computed by the Laerdal software (e.g., the *overallScore* used to determine proficiency) and low-level information for each recorded compression and ventilation event, such as the exact depth of individual compressions (*compDepth* in mm) or the volume of each rescue breath (*ventVolume* in ml).

The scoring algorithm implemented on the SimPad starts everyone with a perfect score and subsequently applies reductions based on non-optimal behavior. **Figure 2** illustrates the level of detail that is available for two example participants at a single measurement point (*CPR.TEST.I.2+FB*). Participant 467 received score reductions because their compressions were too slow and occasionally not deep enough (overall compression score: 80). Their ventilations were near perfect, with slight score reductions for excessive rescue breath volume (overall ventilation score: 96). The combined overall score is 84 for this participant. The other example participant has a combined score of 73, which is a combination of a near-perfect overall compression score (97) but an overall ventilation score of 0 because the rate with which ventilations were administered was too low (and also did not have enough volume).

Data are available for each individual compression event. Therefore, one could dig even deeper to determine, for example, when exactly Participant 467's compressions were too slow and by how much: Did they gradually become slower? Or did they consistently compress at a rate that was just at the threshold and sometimes went under?

Figure 3 illustrates how such questions could be addressed. Shown is the compression frequency associated with each recorded compression during the 1-minute isolated practice with live feedback in Session I (CPR Compressions I). We isolated the participants exhibiting the most and least stable frequencies, where the stability was quantified as the magnitude of the slope of a linear regression line fit to each participant's

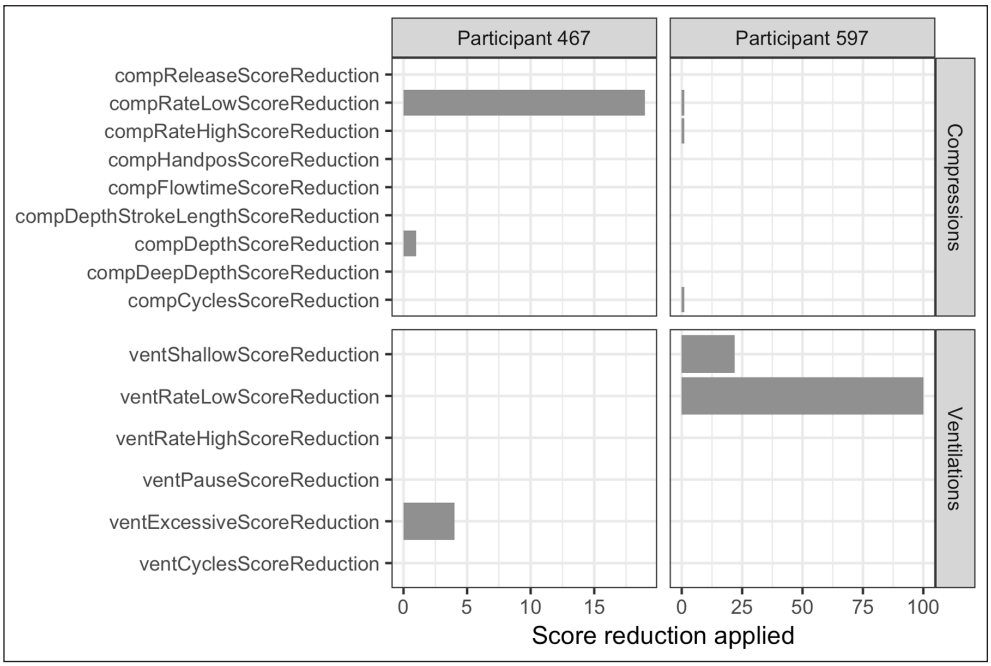


Figure 2: An example of the reductions applied to the scoring for two participants performing a complete run-through of CPR during *CPR TEST I.2+FB*. The labels on the y-axis correspond to entries in the raw data in the column ***type2*** (also see Table 1).

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

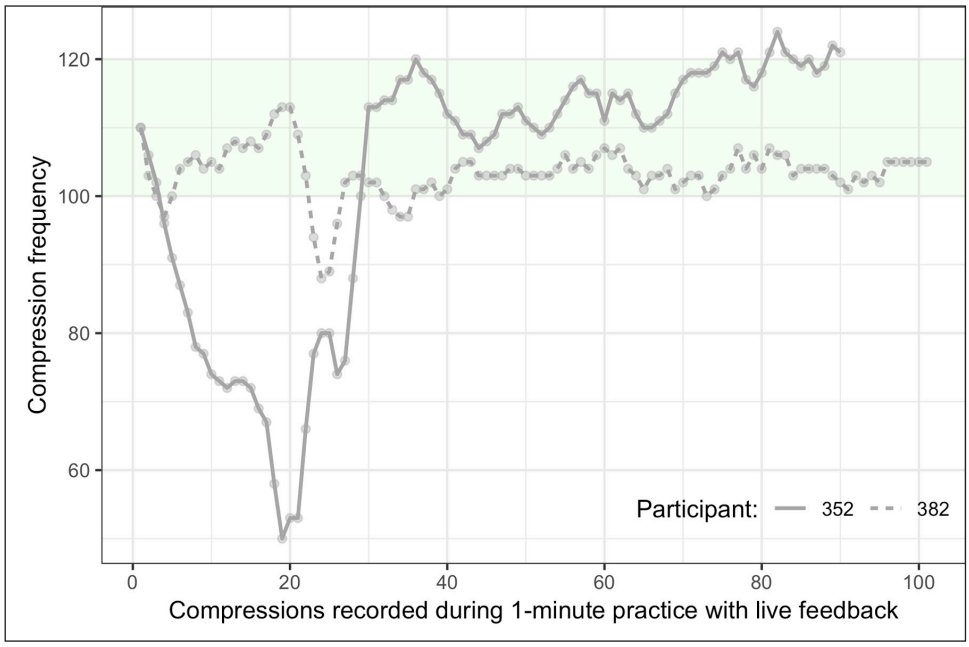


Figure 3: Compression frequencies recorded during CPR Compressions I for the most and least stable participants.

data. Note that this assumes a fairly linear performance profile, which seems to hold in this dataset. More elaborate measures would be needed if, for example, u-shaped performance profiles are expected. As can be seen in the figure, participant 382 shows a very stable pattern with frequencies mostly between 100 and 110, while participant 352's frequencies decrease over the first 20 compressions only to increase for the next 20. Towards the end, 352 is a bit too fast. The compression scores associated with these two performances are 87 and 99, respectively.

This is but a small example of the extreme richness of the CPR data that are available for all participants at various test moments. The supplement provides a more detailed overview and how to extract relevant aspects of the data.

Declarative CPR Performance

The Laerdal manikins cannot measure the steps that should be taken to ensure that initiating CPR is appropriate (i.e., checking breathing, alerting emergency services, etc.). The separate scoring (see *CPR Test I.1* in Methods above for details) provides information on whether the existence and ordering of a step was recalled and completed correctly and scores are available for all *CPR Test* components in **Figure 1** (i.e., complete run-throughs). We refer to these as *declarative* CPR performance to contrast it with the administration of compressions and rescue breaths. The information thus recorded would allow us to, for example, compute which steps were most likely to be forgotten or how long breathing was checked when it was checked (see supplement).

Word Learning

Word-pair study time was fixed at 15 minutes for all participants. Word-pairs were introduced in random order and repetition schedules were personalized to each participant using the adaptive algorithm outlined above. For each participant, time-stamped information for each response includes accuracy, latencies for first and confirming keystrokes, and the estimated model parameters.

Figure 4(A) depicts, for each participant, the proportion of correct responses they gave during word learning (y-axis) and how those relate to the *rates of forgetting* estimated from the adaptive scheduling system (procedure is identical to Sense et al., 2016). The strong negative correlation indicates that participants that made more errors during word learning are estimated to forget the studied material more quickly, as one would expect.

Word Recall

For each participant, information is available for each Word Recall measurement regarding the response they gave on each item. Specifically, the file contains information on the Swahili cue, the correct English translation, and the answer provided to each item — and whether the answer was correct. Participants sub-

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

mitted their Word Recall responses as a batch since all cues were on screen at the same time, which means there are no latencies for individual items.

Each point in **Figure 4(B)** indicates how a participant's proportion of correct responses on $Word_I$ Recall test (x-axis) relates to the *rate of forgetting* estimated during $Word_I$ Learning. The high negative correlation suggests that the rate of forgetting does not only capture individual differences during word learning (cf. **Figure 4(A)**) but also on a subsequent recall test. Additionally, **Figure 4(C)** shows how the Swahili-English paired associates could be ranked in difficulty based on participants' performance on $Word_I$ Recall.

Raven matrices

The Raven matrices were presented sequentially: For each problem, we have information about the response latency, the options that were on screen and how they were arranged, which response option was correct, and which was picked by the participant. Given the 10-minute time limit, not everyone attempted all problems. **Figure 4(D)** shows the relationship between the number of problems attempted and solved by each participant and indicates that many more problems were attempted than solved. This makes sense given the progressively increasing difficulty of the problems.

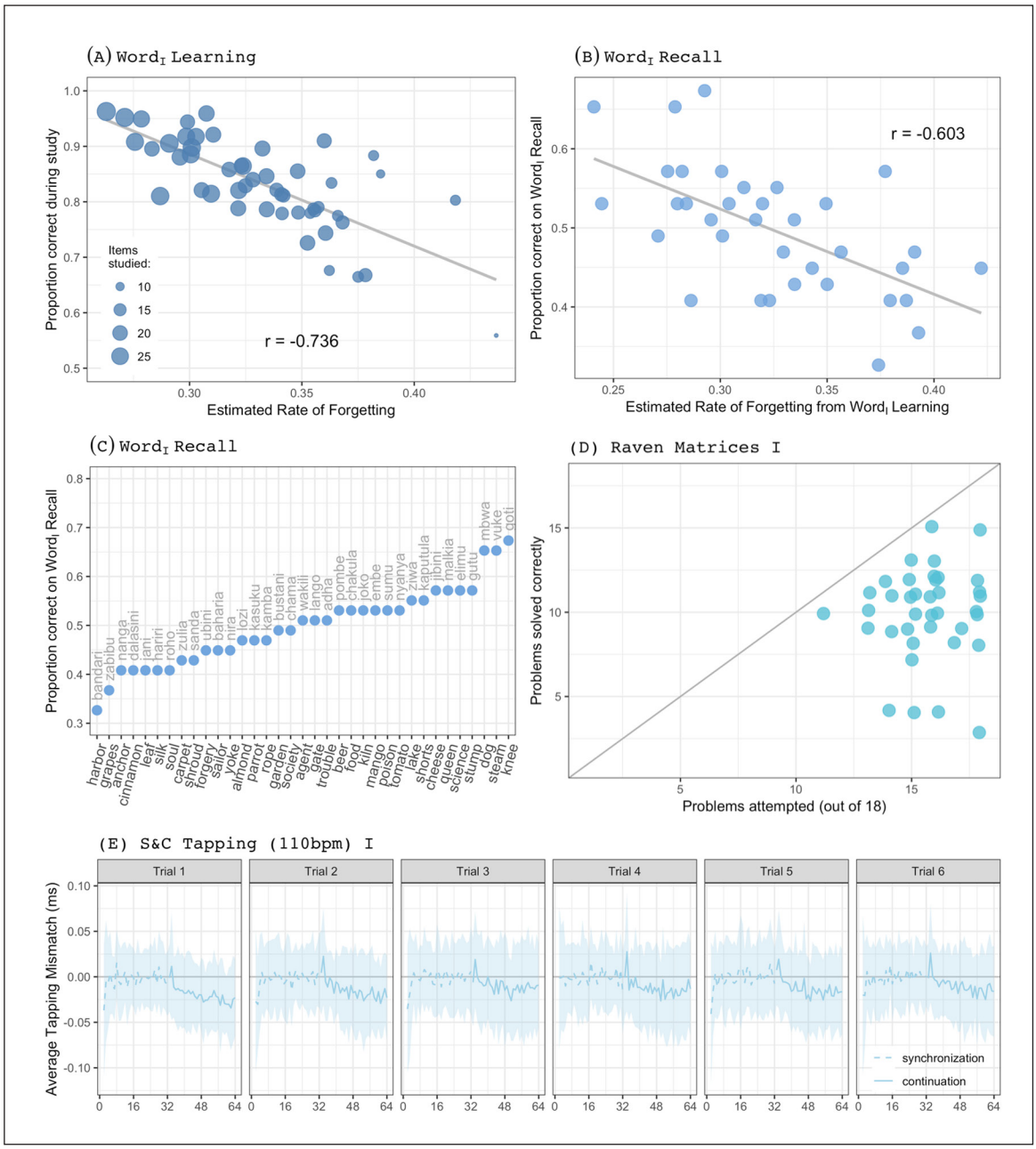


Figure 4: An overview of the type of data that is available for various components. Plot titles refer to labels used in Figure 1. Details are provided in-text. All plots based on data from Session 1 only.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

S&C Tapping

Responses to each individual event in the task are available in the raw data file. Asynchronies are the differences between the onset of a beep (or when it would have continued) and each recorded tap (i.e., tapping mismatch). **Figure 4(E)** shows one possible way to aggregate tapping performance by summarizing the tapping mismatch for the 600ms ISI in Session 1 (i.e., S&C Tapping (110bpm) I). Shown is the average tapping mismatch for each of the 64 events on the 6 trials: The SD is indicated by the ribbon, while the mean mismatch is shown by lines. During the synchronization phase (dashed lines), the mismatch is centered around 0 but as the continuation phase (solid lines) progresses, the average mismatch drifts to negative numbers, suggesting that participants sped up.

Concluding Remarks

In this manuscript, we describe a dataset that we hope will be useful to other researchers. We believe that the type of data presented here are rarely collected: Performance was measured on a number of established lab tasks as well as a complex, real-life task (CPR) across three sessions. Given the richness of these data, we hope other researchers can use them to explore individual differences in performance on a complex, real-life task and traditional lab experiments.

Data Accessibility Statement

All data and scripts are available at https://osf.io/m8bxe/.

Ethics and Consent

All participants gave written informed consent for each session and the study was approved by the Western Institutional Review Board (protocol #20151567) and the local Ethics Committee Psychology (ID: 17017-S-NE(a,b)).

Funding Information

This work was supported by European Office of Aerospace Research and Development (EOARD) grant #11926121.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

Florian Sense and Sarah Maaß authors have shared first authorship to this work.

References

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & **Engle, R. W.** (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*(2), 226–36. DOI: https://doi.org/10.3758/s13421-014-0461-7

Loesche, P., Wiley, J., & **Hasselhorn, M.** (2015). How knowing the rules affects solving the Raven Advanced Progressive Matrices Test. *Intelligence*, *48*, 58–75. DOI: https://doi.org/10.1016/j.intell.2014.10.004

McKenna, S. P., & **Glendon, A. I.** (1985). Occupational first aid training: Decay in cardiopulmonary resuscitation (CPR) skills. *Journal of Occupational Psychology*, *58*, 109–117. DOI: https://doi.org/10.1111/j.2044-8325.1985.tb00186.x

Nelson, T. O., & **Dunlosky, J.** (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory*, *2*(3), 325–335. DOI: https://doi.org/10.1111/j.2044-8325.1985.tb00186.x

Pavlik, P. I., & **Anderson, J. R.** (2008). Using a Model to Compute the Optimal Schedule of Practice. *Journal of Experimental Psychology: Applied*, *14*(2), 101–117. DOI: https://doi.org/10.1037/1076-898X.14.2.101

Perkins, G. D., Handley, A. J., Koster, R. W., Castrén, M., Smyth, M. A., Olasveengen, T., Greif, R., et al. (2015). European Resuscitation Council Guidelines for Resuscitation 2015. Section 2. Adult basic life support and automated external defibrillation. *Resuscitation*, *95*, 81–99. DOI: https://doi.org/10.1016/j.resuscitation.2015.07.015

Raven, J. (2000). The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology*, *41*, 1–48. DOI: https://doi.org/10.1006/cogp.1999.0735

Semjen, A., Schulze, H.-H., & **Vorberg, D.** (2000). Timing precision in continuation and synchronization tapping. *Psychological Research*, *63*, 137–148. DOI: https://doi.org/10.1007/PL00008172

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

Sense, F., Behrens, F., Meijer, R. R., & **van Rijn, H.** (2016). An Individual's Rate of Forgetting Is Stable Over Time but Differs Across Materials. *Topics in Cognitive Science*, *8*(1), 305–321. DOI: https://doi.org/10.1111/tops.12183

Sense, F., Meijer, R. R., & **van Rijn, H.** (2018). Exploration of the Rate of Forgetting as a Domain-Specific Individual Differences Measure. *Frontiers in Education*, *3*(112). DOI: https://doi.org/10.3389/feduc.2018.00112

Soar, J., Nolan, J. P., Böttiger, B. W., Perkins, G. D., Lott, C., Carli, P., Deakin, C. D., et al. (2015). European Resuscitation Council Guidelines for Resuscitation 2015 Section 3. Adult advanced life support. *Resuscitation*, *95*, 100–147. DOI: https://doi.org/10.1016/j.resuscitation.2015.07.016

van Rijn, H., van Maanen, L., & **van Woudenberg, M.** (2009). Passing the Test: Improving Learning Gains by Balancing Spacing and Testing Effects. In *Proceedings of the 9th International Conference on Cognitive Modeling* (pp. 110–115). Manchester, UK.

How to cite this article: Sense, F., Maaß, S., Gluck, K., and van Rijn, H. 2019 Within-Subject Performance on a Real-Life, Complex Task and Traditional Lab Experiments: Measures of Word Learning, Raven Matrices, Tapping, and CPR. *Journal of Cognition,* 2(1): 12, pp. 1–10. DOI: https://doi.org/10.5334/joc.65

Submitted: 30 January 2019 **Accepted:** 05 May 2019 **Published:** 23 May 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

]u[*Journal of Cognition* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited