FFI Norwegian Defence
Research Establishment

21/01631

FFI-RAPPORT

Optimisation of performance-based continuation training

– Study report for ANNC Collaborative Project between NLD and NOR

– NLR Report TR-2021-172

Guro Kristin Svendsen
Jelke van der Pal[1]
Eirik Løhaugen Fjærbu

[1] Royal Netherlands Aerospace Centre (NLR)

Optimisation of performance-based continuation training

– Study report for ANNC Collaborative Project between NLD and NOR

Guro Kristin Svendsen
Jelke van der Pal[1]
Eirik Løhaugen Fjærbu

Norwegian Defence Research Establishment (FFI)
[1]Royal Netherlands Aerospace Centre (NLR)

10 September 2021

Keywords
Modellering
Trening
Flygere
Kampfly
Java (programmeringsspråk)
Optimering

FFI report
21/01631

NLR report
TR-2021-172

Project number
FFI: 1563

Electronic ISBN
978-82-464-3370-7

Approvers

Stian Betten, *Research Manager*
Arne Petter Bartholsen, *Director of Research*

The document is electronically approved and therefore has no handwritten signature.

Copyright

© Norwegian Defence Research Establishment (FFI).
© Royal Netherlands Aerospace Centre (NLR).
The publication may be freely cited where the source is acknowledged.

Summary

This report documents the findings of a collaborative project within the Anglo-Netherlands-Norwegian collaboration (ANNC) between the Royal Netherlands Aerospace Centre (NLR) and the Norwegian Defence Research Establishment (FFI) titled “Optimisation of Performance-Based Continuation Training”. The objective is to explore the potential of optimization of pilot training, with a training program based on competencies and proficiencies (performance-based) instead of currencies (currency-based). This is an exploratory activity, intended to provide recommendations regarding future collaboration in a potential phase 2.

The current continuation training philosophy for the Royal Netherlands Air Force (RNLAF) and the Royal Norwegian Air Force (RNoAF) is that each pilot shall complete a set of training missions within the previous 12 months in order to stay combat ready. These training missions are described in the Annual Training Program (ATP), and have been developed over the years to provide the essential skills and competencies for the average pilot. However, the ATP does not take into consideration the fact that pilots are different individuals, and that some pilots need more training in some areas, and less in other areas. If the training program could be tailored to the individual pilot needs, the pilot skill level would increase – and the hours spent performing unnecessary training missions could be reduced, increasing pilot performance, and decreasing the resources spent on pilot training.

NLR is working on a competency-oriented, performance-based training (PBT) approach for defining continuation training programs. PBT allows for the training to be adapted to individual needs, but requires a flexible training system which makes it more challenging to schedule pilots to training missions. FFI has developed a simulation tool for analysis of the feasibility of executing a training program – “Training for F-35 Pilots” (TREFF). TREFF simulates the execution of a training program given a set of resources and constraints, taking stochastic events and scheduling into account. Combining the two approaches would provide a possibility for NLR and the RNLAF to test the feasibility of proposed training programs, and for FFI and RNoAF to justify the composition of the continuation training program.

The report gives an introduction to the NLR and FFI training analysis approaches, and addresses some possibilities on how the FFI tools and the NLR products can be combined. In order to explore the options for interfacing products and tools, a test case is defined. We also outline a PBT concept, and provide a roadmap for the implementation of a PBT approach.

Sammendrag

Denne rapporten dokumenterer funnene i en samarbeidsaktivitet mellom the Royal Netherlands Aerospace Centre (NLR) og Forsvarets forskningsinstitutt (FFI) med navn «Optimisation of Performance-Based Continuation Training». Samarbeidet foregikk innenfor rammen av Anglo-Netherlands-Norwegian collaboration (ANNC). Målet med aktiviteten er å utforske potensialet for optimalisering av trening av piloter, med et treningsprogram basert på ferdigheter og kompetanse (ytelsesbasert) i stedet for frekvensbasert. Aktiviteten skal gi anbefalinger om videre samarbeid i en potensiell fase 2.

Dagens treningsfilosofi for kontinuitetstrening i begge lands luftforsvar (RNLAF i Nederland og RNoAF i Norge) er at hver pilot skal ha gjennomført et sett med treningsoppdrag de siste tolv månedene for å holde seg kampklar. Disse treningsoppdragene er definert i det årlige treningsprogrammet, og har blitt utviklet gjennom årene for å gi de viktigste ferdighetene og kunnskapen som en gjennomsnittspilot trenger. Men det årlige treningsprogrammet tar ikke med i betraktningen at pilotene er forskjellige individer, og at noen av dem trenger mer trening på noen områder og mindre på andre områder. Hvis treningsprogrammet kunne blitt skreddersydd etter pilotenes individuelle behov, ville ferdighetsnivået øke – dessuten vil det kunne redusere antallet timer brukt på unødvendig trening, som vil kunne føre til bedre pilotprestasjoner og bedre utnyttelse av ressurser i forbindelse med trening.

NLR arbeider med en kompetanseorientert, ytelsesbasert tilnærming til treningsprogrammer for kontinuitetstrening. Ferdighetsbasert trening gjør det mulig å tilpasse treningen til individuelle behov, men krever et fleksibelt treningssystem, noe som gjør det vanskeligere å planlegge for piloter til treningsoppdrag. FFI har utviklet et simuleringsverktøy, Trening for F-35-flygere (TREFF), for å analysere gjennomførbarheten til treningsprogram. TREFF simulerer gjennomføringen av et treningsprogram, gitt et sett med ressurser og føringer, og tar stokastiske hendelser og allokering av piloter med i betraktningen. Å kombinere de to fremgangsmåtene vil gjøre det mulig for NLR og RNLAF å teste gjennomførbarheten av deres foreslåtte treningsprogram, og for FFI og RNoAF å kunne underbygge innholdet i treningsprogrammet.

Rapporten gir en introduksjon til FFIs og NLRs fremgangsmåter for treningsanalyser, og beskriver hvordan FFIs simuleringsverktøy og NLRs forskningsresultater innen kompetansebasert trening kan kombineres. For å utforske dette definerer vi et test-case. Rapporten skisserer også et konsept for ytelsesbasert trening, og gir et veikart for implementering av ytelsesbasert trening.

Contents

1 Concept and methods

1.1 Performance-Based Continuation Training

There may be several perspectives on Performance-Based Training (PBT) for maintaining proficiency of qualified personnel[1]. A common and central feature is the departure from a currency-based training approach, in which the professional, in this report the fighter pilot, would require refresher training with a certain frequency. This could, for example, be twice a year for a specific training objective, which may relate to a task, a condition, or specific skills or knowledge. PBT aims to use actual performances to predict the optimal amount and timing of the next training. For refresher training (which is the major part of Continuation Training for fighter pilots), the performances relate to proficiency levels, therefore PBT might also stand for proficiency-based training.

Another feature of PBT perspectives is its data-driven base, the notion that proficiency should be measured as objectively as possible. If objective data cannot be obtained, qualitative data (instructor or self-ratings) should be sufficiently discriminative (not just pass/fail), and the rating process should be standardized and calibrated.

In civil aviation this objectification of instructor ratings is part of what they label as Evidence Based Training (EBT) and which is formalized in ICAO (International Civil Aviation Organization) guidelines and EASA (European Union Aviation Safety Agency) regulations. EBT for civil aviation is a training philosophy that has been in development since 2007 and has been in use with several airlines for a few years now. Civil-EBT is rooted in the concepts of resilient operations, and aims to develop pilots in several core competencies, instead of training a narrow, familiar and repetitive set of tasks. By developing competencies, a pilot is effectively prepared to manage safety and performance in many different situations, not only those explicitly trained. Civil-EBT features the combination of evidence (data) driven training and competency based training. Civil-EBT generates competency data, and uses this data to adapt the operator's recurrent training to focus on actual competency training needs. In this way, Civil-EBT provides a framework to continuously improve pilot core competencies.

Competency-Based Training (CBT) represents a first step towards PBT. CBT seeks to focus on the competencies, skills, and knowledge underlying the missions and tasks to train, as well as on the conditions (complexity factors) that make task performance more or less difficult. For continuation training, experienced or estimated skill decay on competencies may drive the training schedule. As such, CBT may generate refresher training with predetermined intervals for each competency. Although the intervals can be based on experience, and therefore empirically grounded, they do not reflect individual differences in training needs. In that

---

[1] In military aviation, proficiency is often seen as related to (instrument) flight specific skills which are subject to regular proficiency tests for regulatory safety reasons. In PBT, all skills, flight as well as tactical, are to be monitored on proficiency for training efficiency reasons next to safety reasons.

respect, CBT does not differ from currency-based training, which focuses more on missions and tasks.

To our perspective Civil-EBT is a combination of CBT and PBT including evidence from operational performance, where the PBT component strongly depends on instructor ratings, although improved and standardized. As such, EBT is a step towards full-scale, objective PBT. In our framework of training concepts, we distinguish EBT from PBT. EBT relates then to the results of measured performance in operations (evidence that training results transfer to the operational world, i.e. actual (deployed) missions) are being used to improve or optimize training. In training practice however, EBT should be used together with PBT.

With sufficient amount of performance data from training and/or operations, the statistical power (or its equivalent in AI techniques) of predictions on (fading) performances may be strong enough to be reliable, even on a personal level. Such performance/evidence-based personalized training will require highly flexible training schemes.

A final development of PBT may be a real-time adjustment of the level of difficulty or injections of specific events during a (simulated) training session as based on the performance and/or the effort a pilot puts into the performance (cognitive load) within the training session. Such real time, automated personalized adaptation would be the ultimate end state of PBT.

Figure 1.1 A family concept of training theories.

While any of the above-mentioned training theories may be applied without a competency-based perspective or using performance data – obviously, good instructors achieve adaptive personalized training on their own merit, we anticipate great benefit of a systematic build-up of these applications. This way, a family concept of training theories is formed with competency-based training as a start, to which performance-based models are added as well as evidence-based models; see Figure 1.1. Only then, a firm basis will be obtained for automated personalized training within a larger training plan or automated adaptive training within a single training session.

In an ideal world, where personalized training has been optimized, there is little need for adaptive training within a training session. The most appropriate scenarios to restore loss of performance should have been selected at the most optimal timing. Some adaptation may still be effective in case of uncontrolled conditions, such as personal fitness, or distractions. The system may detect unforeseen loss of performance and lower the difficulty level for example. The focus therefore should be given to optimize personalized training. In this project we explore techniques and options for optimizing continuation training by performance-based and if possible personalized approaches.

Skill decay/retention

For continuation training, the primary data to collect relates to the loss of skill that can be expected after a period of non-use, also known as skill decay. In recent years, the tendency to rephrase the notion in a more positive way is increasing; the flipside of the coin is called the retention of skill. The period of non-use is often labelled as the “retention interval”. Abundance of empirical data on skill decay is available for elementary skills or knowledge items. It has been found that recently acquired skills and knowledge decay according to a power law curve. The slope of the curve differs between tasks (and associated skills), for example the psychomotor skills for physical tasks retain much better than the cognitive skills for procedural tasks. The approach and quality of training also has an effect on retention. Distribution of practice over time (spaced practice as opposed to massed practice) is very beneficial for the retention of skills. There are personal factors as well. Cognitive ability and motivation will influence the level of retention. Several more influencing factors have been reported. Full overview can be found in [1, 2] or in the upcoming final report of NATO RTG HFM-292 [3].

However, little is known about interaction effects of these factors and how retention works for more complex skill sets, where several elementary skills have been integrated to perform a complex task such as a basic fighter maneuver. Skill decay for more experienced pilots on such complex skills may not decay according to a power law curve, but have incidentally been found to follow a much slower S-curve decay pattern [4], which is line with our personal and professional experiences. Also, some knowledge can be remembered for ever, and certain skills will remain intact for a very long time as well (such as riding a bicycle). The S-curve for such skills may be stretched very far.

The process of change in the retention curve as expertise is built, and how this differs between persons and tasks, can only be understood by collecting more data with more details than we do today. Short-term experiments will not be sufficient.

Predictive modeling

With initial assumptions on decay functions, a basic predictive model may be formulated. The model can learn from new performance data, which allows for more specific decay functions, while influencing factors may be added gradually to increase the predictive power. The model may apply to a group initially and grow into personalized versions as predictive power allows it.

1.2 Design elements for Performance-Based Training

NLR's approach to competency-based training is based on Van Merrienboer's Hierarchical Skills Analysis [5] and elements from the (U.S.) Air-Force Research Laboratories (AFRL) Mission Essential Competencies (MEC) approach to training [6]. There are two broad perspectives on competencies: a training perspective and a Human Resource Management (HRM) perspective. In the training perspective, a competency is an integration of skills, knowledge and attitudes. In this hierarchical approach, a competency may be built up from more detailed constituent competencies in a hierarchical way, where ultimately elementary skills, such as reading or arithmetic may be found at the bottom of the hierarchy. For training purposes, the levels that are already mastered at the start of the training do not need to be expressed. The HRM perspective is more concerned about the high-level, non-technical skills or capabilities, such as communication, creativity, flexibility. These competencies tend to be more generic and transferrable between jobs and functions, and are helpful in career planning. Both perspectives are relevant and used in NLRs approach to CBT (there labelled as supportive competencies, as they often do not fit in the hierarchy of task-oriented competencies and skills). Please note that certain organizations, such as EASA competency-based training guides for civil aviation, primarily use the HRM perspective when referring to competencies.

For qualification training, all competencies need to be mastered and trained. A whole-task sequencing principle is advised for that, supported by part-task practice where needed. For continuation training the NLR approach suggest a stronger focus on part tasks that require (more specific) competencies susceptible for skill decay, provided that 1) the pilot is fully proficient and capable to start with, and 2) full missions are practiced on a regular basis in an annual program. During the larger scale full missions, the integrative and collaborative aspects of the competency profile are challenged. Proficiency issues may be found on the collaborative, integrative level, which may (or may not) in turn be caused to decay on more specific skills. These more specific competencies, on a lower level in the skills hierarchy, are suggested to be trained in a highly personalized training regime, using simulators or small-scale live training that focus on part tasks, such as specific Tactics, Techniques, and Procedures (TTPs).

Indicators for skill decay may be standard performance measures (measures of effectiveness or measures of performance), instructor ratings (if any for Continuation Training sorties), and self-reports, although the most useful indicators on a more personal level are assumed to be more specific and may not yet have been identified (this may require R&D on eliciting knowledge from experienced and excelling instructors as well as more in-depth analysis of (tactical) Flight Data Monitoring and simulator data).

Tasks performed under different circumstances may require competencies that are basically different, although we use the same skill label for them. For example, certain weather conditions obviously make flying or tactical tasks more difficult and the competencies need to be practiced and maintained in each of these conditions. We refer to circumstances that make the task more or less difficult as Complexity Factors. Examples of complexity factors include the mobility of the targets, lighting conditions and technical issues with the aircraft. During qualification training, complexity factors can be used to structure the training sequence; for continuation training, retention needs to be ensured for the competencies on each complexity factor.

1.3 Simulation-based analysis of training

FFI uses discrete-event simulations (DES) [7] to support analyses for the Norwegian Armed Forces in a range of areas. In particular, we have used DES to support decisions regarding design or redesign of training systems. Obtaining sufficient training depends on best use of limited resources. Time is often a key aspect. From the perspective of an individual trainee, he or she should optimally receive training according to a certain progression. This will require access to suitable training media and instructors. In addition, there is often the need to train together with other trainees, and possibly also with someone acting as the enemy forces. The training system consists of instructors, supporting actors, training media, a training program and intended users. This system must be designed to give sufficient training with reasonable use of resources. In our analyses, we also include the stochastic element, to see how vulnerable the system is for unwanted effects like grounded aircraft or difficult weather.

The TREFF analysis tool [8] is one example which was used to advice regarding combat aircraft pilot training. Training for combat aircraft pilots is a complex task from a resource allocation perspective. The training requires a set of training media, which could be for example aircraft or high-fidelity simulators. Further, the pilots train in formations. Optimally, the mission under training should fit the needs of all trainees. In reality, this is often a difficult scheduling issue, especially for squadrons with an emphasis on training tactics in larger formations. TREFF was designed and used to evaluate the ability of alternative training systems to provide sufficient training. A currency-based annual training program constituted the requirements for pilot training. A pilot conducting all missions in accordance with this training program should in principle be combat ready. However, the final assessment of combat readiness will be based on operative judgement from a leader. Training in accordance with the training program should still be a good indication of the combat readiness. The main question for TREFF to address was: can the training required by an annual training program be completed within one year, and if so, how many days margin is there?

TREFF simulates the training sorties flown by the different pilots, based on stochastic factors like weather and the availability of aircraft and simulators. Training is scheduled according to a training program consisting of a set of missions with associated annual currency requirements. In addition, the pilots must contribute as supporting actors by manning red-air formations (i.e. acting as the enemy). The missions included in the training program are not defined in detail, but refer to general types of missions such as Basic Fighter Maneuvers, Defensive Counter Air and Close Air Support. When conducting these missions, there may be differences from sortie to sortie. Different sorties of the same mission may differ in the exact events included, as events are not specified explicitly in the TREFF training program. For each mission, the training program in TREFF will specify a target number of aircraft in the formations on both blue (friendly) and red side, but there is some flexibility, so that it is possible to train the mission with fewer aircraft if necessary. Time constraints are included in the model using a set of generic daily training schedules. These provide constraints as to how training slots can be combined for a given pilot.

As the main question of TREFF is whether the training system has sufficient capacity to allow the pilot to complete a yearly training program within time, TREFF has less focus on the details of how this is achieved. The scheduler therefore attempts to complete as much training as possible as fast as possible. In reality one would rather distribute the training as evenly as possibly over the year, and also schedule the different missions in an order supporting the best training progression for the pilot. In addition, TREFF does not distinguish between pilots based on qualification and experience. Constraints on which pilots can fill which seats in a formation is therefore not included, and the training requirements are based on an average between requirements for experienced and unexperienced pilots. A high-level generic sketch of the TREFF model is shown in Figure 1.2.

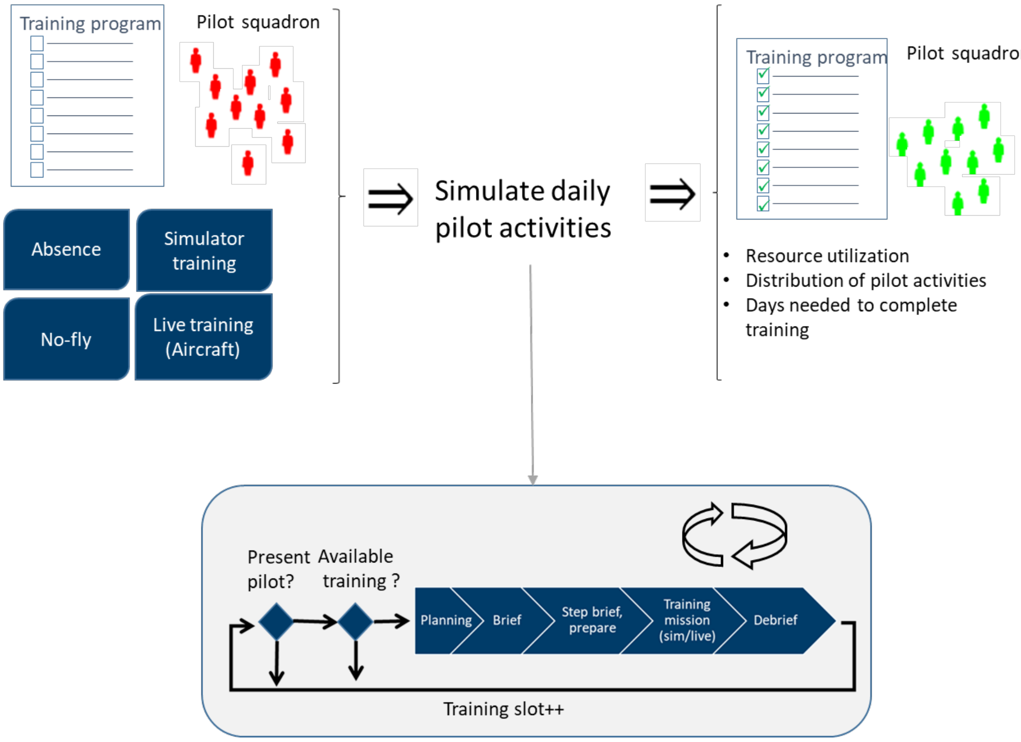


Figure 1.2 High-level generic sketch of TREFF. An annual currency-based training program specify the training requirements for a squadron of pilots. The focus of the model is on the activity of these pilots. The main categories of activities are absence (illness, vacation ...), training in simulator or aircraft, and no-fly activities like administrative issues and theoretical lessons. TREFF simulates daily pilot activities, taking stochastic factors into account. The simulation provides insight into resource utilization, distribution of pilot activities, and days needed to complete the training program.

In order to answer more detailed questions regarding how to conduct the training, a new simulation tool was developed – TREFF2 [9]. TREFF 2 includes a more sophisticated scheduler, trying to distribute the training evenly. The model also distinguishes between different categories of pilots. As the work described in this report describes a first step towards evaluating the potential of joining the approaches of FFI and NLR, we have opted to use the original TREFF in this work. Since it provides less details, it is more lightweight, will be easier to use, and provides a more rapid overview of the potential. It might be an option to replace TREFF with TREFF2 in later stages.

2 Joint possibilities

2.1 Future research questions

A joint approach could be aimed at investigating the following three research questions:

A. How can competency profiles be used to create a training program?

In order to reach this goal, the competency profiles should be clearly defined. Several ambition levels are possible, from a squadron profile, through experience level profiles to individual profiles. A means of periodic measurement of competency is needed in order to realize the highest ambition of such personalized training.

There are several approaches for the measuring performances associated to the identified competencies:

- Instructor rating of competency measures: 'Civil EBT style' standardized instructor ratings. For that, the identified competencies need to be worked out in behaviors that are observable to instructors or the person responsible for ratings. Training of instructors is required to standardize the ratings.

- Data-assisted rating of mission measures: Mission-related performance measures and certain competency performances may be measured objectively. One way to achieve this is by using a tool, such as the AFRL Performance Evaluation Tracking System (PETS)

- Self-ratings: Pilots rate themselves after each (sim) sortie. It is not feasible to rate all competency elements in detail. Perhaps only marking elements to which the pilot felt 1) some level of skill fade causing them to actually make errors/take too much time, or 2) really at ease to perform / had much spare mental capacity left.

Once the competency profiles are defined, they must be translated into some set of requirements for the training and a training program. Such a training program will generally be much less specific than a currency-based training program.

B. *How will the flexibility of personalized training affect the feasibility of the scheduling?*

Personal needs can be expected to differ considerably among pilots as a result of different experiences and different personal retention curves for the competencies. Training will need to be scheduled with a basis in missions with certain formation sizes, and there will be a mutual dependency among the pilots in the formation. The same mission will generate different training value for each pilot, and they may be focusing on different competencies in the same training mission. At the same time, a competency can – of course – be applied in missions even when it does not need training yet. How to address and process such experiences is not yet fully conceptualized. On one hand, there is an expected advantage with increased flexibility that there are many ways to reach a certain competency level. On the other hand, it may be more challenging to identify training missions that match with the different competency profiles.

Feasibility analyses could be used to investigate how these factors affect the training once the PBT-concept is more materialized for a certain squadron.

C. *How can competency profiles be used to create a scheduling tool?*

Competency profiles can be specified at various levels of specificity. Their primary use is to provide a high-level perspective of the most critical (groups) of competencies. However, this may include some very specific skillsets. Currently, it is not fully known which specific skills that needs to be included in such a competency profile. A scheduling tool must balance available resources and time constraints with the requirements from these competency profiles and the associated overall operational requirements.

2.2 First steps

In this exploratory work, the following questions have been investigated further:

I. *How can the NLR approach to CBT be used to provide input for TREFF?*

This includes a mapping from PBT-parameters (competencies, complexity factors, retention intervals) to missions and frequency requirements. A simple linear retention model will be used as an illustration, and output of estimated competency-level for the pilots will be added to TREFF.

Simulation results can be used to identify problematic aspects of the training program, like competencies that it is difficult to provide sufficient training for.

This will be a first step to investigate research question A) using a simulation model like TREFF.

II. *What are the possible challenges with the flexibility of personalized training, and what are the relevant metrics for a scheduling tool?*

A full analysis of this will require a scheduling algorithm using the competency profiles. In order to gain more insight into possible challenges and recommended metrics for the scheduler, we will use the traditional scheduler, and show the resulting variation in competencies as an output. Personal variation will be simulated using a simple stochastic model with variation in learning outcome and in the personal retention intervals.

This will be a first step to investigate B), and also provide some insight into possible metrics for a scheduling algorithm C).

3 Test case

The test case was created in order to explore future joint possibilities of the training analysis approaches of NLR and FFI. The starting point of the test case was simplified sets of training missions, competencies, and complexity factors, in accordance with NLR's approach to competency based training. The missions, competencies, and complexity factors, as well as the baseline skill-retention model, are described in section 3.1.

Based on this input, we created a more specific training program that is suitable as input to TREFF. This implies creating a currency-based training program compatible with the existing scheduling system in TREFF. This training program, and the method we used to formulate it, is described in section 3.2. Section 3.3 describes the remaining training system characteristics like trainees, training media and available training slots.

The test case was simulated in TREFF, and functionality was added to the simulation tool in order to track and analyze data related to PBT. Here, we include stochastic variations to the learning outcome and the individual retention intervals. The simulation results are summarized in 3.4, and discussed in 3.5.

3.1 Competency-based training requirements

The missions, competencies, complexity factors, and retention intervals are the main constituents of the competency-based training approach. We use abbreviations for the competencies and complexity factors; the full names are listed in appendix A. We emphasize that the focus of this work is on the method used, and that the data provided in this specific test case is not intended to represent a realistic training system.

We use a numbering scheme for the missions in the test case; see Table 3.1. We group the missions into two types: part-task missions (PT) focused on specific tasks and tactical missions (TM) that more closely resemble full (deployed) missions. In real squadrons, there are many aspects of each mission that can be adapted to give the most suitable training for everyone involved. In the test case, one can modify the missions by varying which complexity factors are included and the number of aircraft involved. Figure 3.1 shows the requirements for the numbers of pilots in the red and blue formations, and how much the numbers can vary.

Mission Number	Mission Name
PT1	Emergency Procedures
PT2	Basic Fighter Maneuvers
PT3	Air Combat Maneuvering
PT4	Tactical Intercept
PT5	Shooting Range
PT6	Surface Attack Tactics
PT7	Surface Attack Tactics (with Opposing aircraft)
TM1	Defensive Counter Air
TM2	Offensive Counter Air (Small Formation)
TM3	Offensive Counter Air (Large Formation)
TM4	Close Air Support
TM5	Suppression of Enemy Air Defences
TM6	Suppression of Enemy Air Defences (with Opposing aircraft)

Table 3.1 Full names of the missions used in the test case and the numbering scheme.

The test case includes 20 relatively high-level competencies. Some are very broad, involving a wide range of constituent competencies, whereas others are somewhat more specific. The more specific competencies require the most frequent training, so they have a significant impact on the total need for training. For each mission, we have created a mapping to specify which of the competencies can be trained. In most cases, the competencies are inherent to the mission and therefore always included, but some of the competencies are optional; see Figure 3.2.

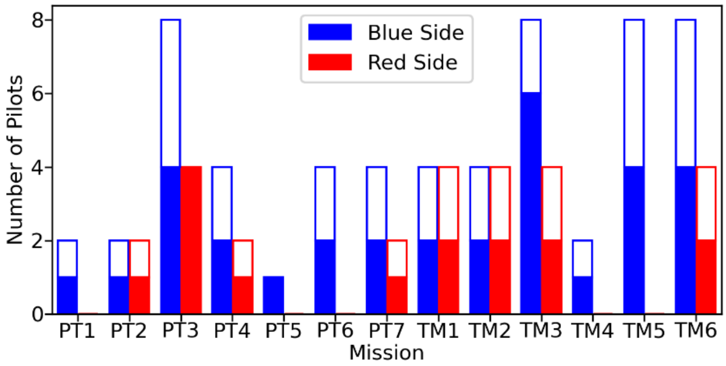


Figure 3.1 Required number of pilots in each mission, on the blue and red sides. The solid bars indicate the minimum required numbers, and the outlines indicate the ideal maximum number.

We use a simplified linear skill-decay model for the competencies. Here, the skill level is quantified by a proficiency P, where P=1 is the maximum. The pilots should retain a proficiency higher than a threshold, here set to P=0.9, to ensure combat readiness. If the pilots do not carry out any training, the proficiencies will decay from P=1 to P=0.9 over a period of R weeks, where R is the retention interval. Each time the pilots carry out a training mission, they gain proficiency for all competencies relevant to that mission. In this test case, we assume that all proficiencies are increased by the same amount after a training sortie. We refer to that amount as the training outcome T. Based on this, TREFF was extended to calculate the proficiencies during the simulation. Stochastic variables were used for both the retention intervals and training outcome.

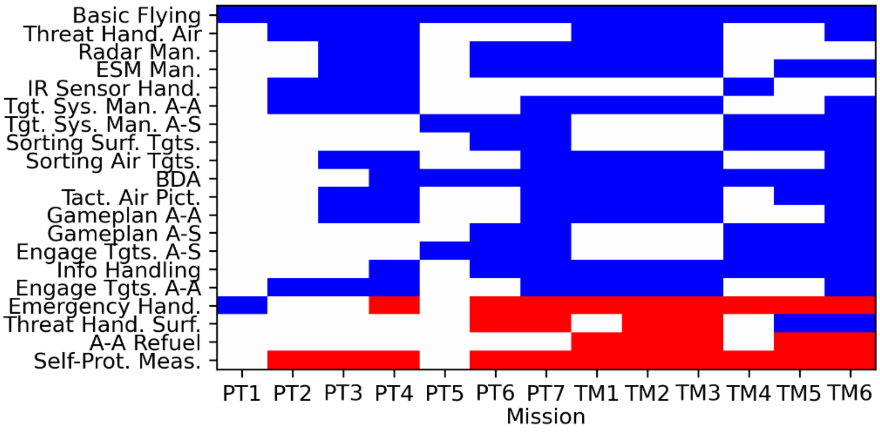


Figure 3.2 Mapping showing which competencies the pilots can train for during each mission. The blue blocks indicate that the competency is inherent to the mission, and red blocks indicate that the competency is optional. For example, a mission may or may not involve use of self-protection, and this affects the training outcome.

To model variations in the quality of the training sorties, we draw T from a probability distribution; see Figure 3.3. We draw a new value of T each time a training sortie is carried out, but use the same value of T for all the pilots that are part of the blue formation and all the proficiencies that are trained.

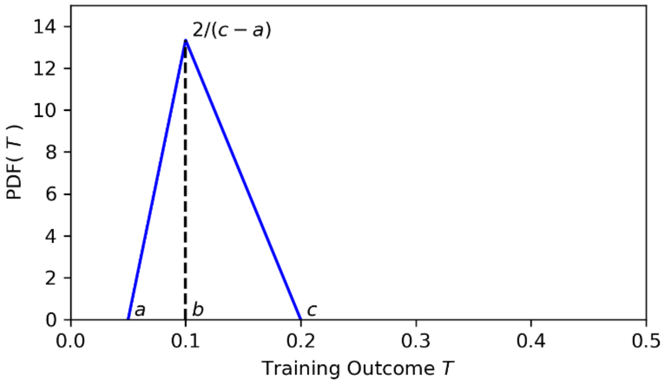


Figure 3.3 Probability distribution of the training outcome T. When a pilot completes a training sortie, his/her proficiencies (for the relevant competencies) are increased by this amount, or up to the maximum proficiency of 1. We use a triangular distribution with minimum value a=0.05, peak b=0.1, and maximum value c=0.2.

To model individual variations, we draw retention intervals R from a Gaussian distribution; see Figure 3.4. We draw independent intervals for each pilot and each competency at the start of the simulation. The mean value and standard deviation of the distribution depends on the competency, but is the same for all pilots. The distribution parameters for all competencies are given in Table 3.2.

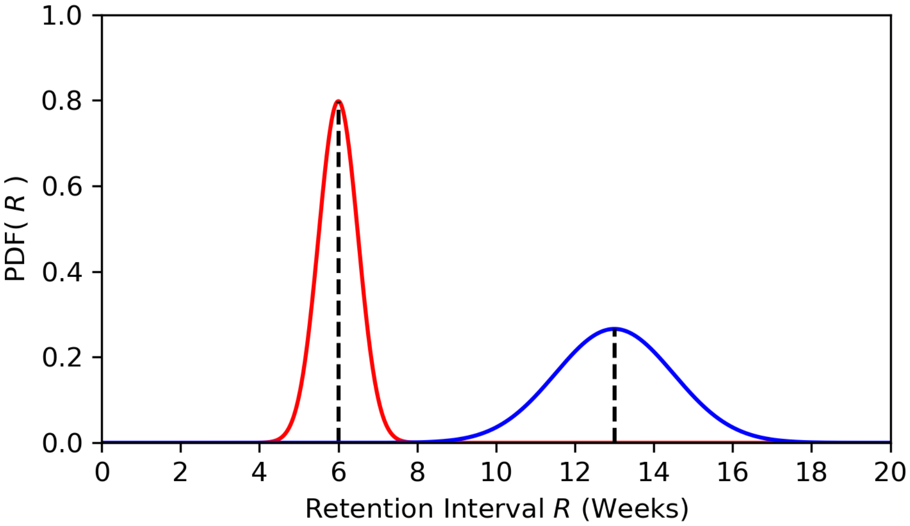


Figure 3.4 Examples of probability distributions for the retention. The retention intervals have Gaussian distributions where the mean values and standard deviations depend on the competency.

Competency	Mean (w)	SD (w)	Competency	Mean (w)	SD (w)
Basic Flying	13	1.5	Tact. Air Pict.	13	1.5
Threat Hand. Air	8	1	Gameplan A-A	26	3
Radar Man.	6	0.5	Gameplan A-S	26	3
ESM Man.	6	0.5	Engage Tgts. A-S	13	1.5
IR Sensor Hand.	6	0.5	Info Handling	13	1.5
Tgt. Sys. Man. A-A	6	0.5	Engage Tgts. A-A	13	1.5
Tgt. Sys. Man. A-S	6	0.5	Emergency Hand.	8	1
Sorting Surf. Tgts.	13	1.5	Threat Hand. Surf.	8	1
Sorting Air Tgts.	13	1.5	A-A Refuel	52	6
BDA	13	1.5	Self-Prot. Meas.	8	1

Table 3.2 Expectation values (mean) and standard deviations (SD) for the retention intervals of the competencies, given in units of weeks (w). We draw the intervals for each pilot and each competency at the start of the simulation experiment. They are drawn from Gaussian probability distributions with the given parameters.

The missions can be modified with respect to complexity factors, which represent conditions that affect the pilots during the missions. Figure 3.5 shows which complexity factors are relevant to each mission. The complexity factors allow more variation in the training, which may improve the pilot proficiency. In contrast to the competencies, the complexity factors are optional in most cases, but a few are also inherent in certain missions. We refer to each complexity factor as having an on-state and an off-state, where the on-state is the more complex one. The retention intervals refer to the on-state. Note that in this model, the complexity factors are treated in the same way as optional competencies, but the underlying rationale to include them is different. The competencies represent tasks that the pilots must master in order to be combat ready, and therefore have to practice sufficiently often. The complexity factors represent different conditions that the pilots should be able to handle, but the training requirements for the complexity factors are less strict.

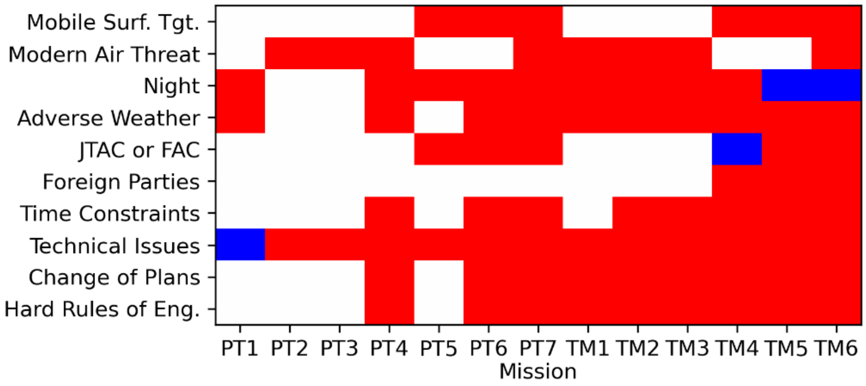


Figure 3.5 *Mapping of complexity factors and missions. The red blocks indicate complexity factors that are optional to a mission, and can be added or removed if necessary. The blue blocks indicate that the complexity factor is inherent in the mission and cannot be removed. White blocks indicate that the complexity factor is not applicable.*

We include retention intervals for each complexity factor and each mission. The mission retention intervals reflect the retention of context-specific competencies within a specific mission, and ensures that the pilots practice all the missions. Similarly, the retention intervals for the complexity factors ensure that several different variants are practiced within the missions. However, we must take care to balance these retention intervals against those of the competencies. For example, if the retention intervals for the missions are too short, the competency requirements are always satisfied as long as the mission requirements are fulfilled, so the training program effectively becomes currency-based. The retention intervals for the missions and complexity factors are given in Table 3.3.

Mission	**Retention Interval (w)**	**Complexity Factor**	**Retention Interval (w)**
PT1	8	Mobile Surf. Tgt.	13
PT2	16	Modern Air Threat	20
PT3	20	Night	8
PT4	26	Adverse Weather	20
PT5	52	JTAC or FAC	8
PT6	26	Foreign Parties	20
PT7	26	Time Constraints	13
TM1	26	Technical Issues	13
TM2	26	Change of Plans	13
TM3	26	Hard Rules of Eng.	13
TM4	26		
TM5	26		
TM6	26		

Table 3.3 *Retention intervals for the missions and complexity factors, in weeks (w).*

3.2 Currency-based training program

The competency-based training requirements of the previous section are defined in terms of (generic) missions, competencies, complexity factors, and retention intervals. By contrast, the currency-based training programs used in TREFF for this test case are based on lists of more specific variants of the missions. Each item in a TREFF training program represents a mission, and specifies the required formation sizes and the number of required repetitions in aircraft and in simulator. For the test-case, we have also extended the training program to include information about the competencies and complexity factors for each item. We use the formation sizes shown in Figure 3.1.

The original TREFF does not focus on the order of the training events, but for this test-case, we also wish to consider the evolution of the training over the year. To spread out the training over the year, we make use of the fact that the TREFF scheduler tries to carry out the missions in the order they are listed in the training program. To give as much variation as possible, we include only one repetition (either live or simulator) for each item in the list. However, some items are identical, corresponding to doing the same mission variant at different times of the year. Note that a scheduler based on competencies should be developed for a full analysis of a realistic training program.

The currency-based training program that we use for this project therefore consists of a list of specific training sorties where the mission type, competencies, complexity factors, and training medium are all specified. We have developed a tool called COmpetency-Mission Frequency Optimizer for Readiness Training (COMFORT) to create such a training program that fulfills the competency-based training requirements [10]. This tool takes competency-based training requirements on the form given in section 3.1 as input, as well as a cost function. COMFORT determines the desired sequence and frequency of missions by solving a constrained optimization problem (COP), which is achieved using OR-Tools [11]. Using COMFORT, we minimize the total sum of costs for all the missions in the training program, subject to the requirements described in the following paragraphs.

We consider a period of half a year, and divide by the expectation values of the retention intervals to get the minimum number of repetitions for each competency. Thus, all pilots get the same training program, whereas the retention intervals in the simulation itself vary between pilots. Additionally, we require that at least half of the sorties are live, and that Air-to-Air Refueling is only possible in live training. We then define a cost for each mission, complexity factor, and optional competency; see Figure 3.6. We also add an additional cost (of 2 units) for aircraft training relative to simulator training.

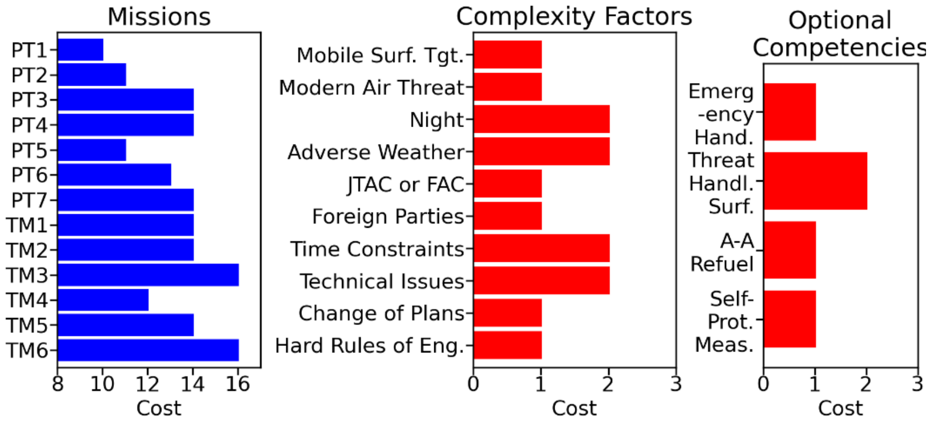


Figure 3.6 COMFORT will minimize a cost function, where the cost of a specific mission is given by the sum of the inherent mission cost and the cost of all the complexity factors and optional competencies.

The minimum number of repetitions that we derived from the retention intervals are only sufficient in the ideal case when all competencies are trained exactly at the end of their retention intervals. This is not achievable in practice, since it will occur that multiple different pilots each have to practice a different set of competencies on the same day. There is only one time slot for live training and two for simulator training each day, and the number of aircraft and simulators available is limited. To account for this, we scale all the competency requirements by a redundancy factor of $C = 2.5$[2]. This redundancy factor ensures that the pilots perform additional training within each competency beyond what they would need in the ideal case. However, as each mission includes several competencies, and because we include repetition requirements for the missions and complexity factors, the relative increase in the number of sorties is significantly smaller than the redundancy factor.

The amount of redundancy needed depends on the characteristics of the squadron (such as the number of pilots and aircraft). One way to determine C is to formulate requirements for how much of the time the pilots should be combat ready, execute simulation experiments with different values of C, and choose the smallest value that satisfies the requirements. Out of the three types of retention intervals: missions, competencies, and complexity factors, the competencies have the highest priority. Therefore, we include a redundancy factor for the competencies only.

The training program outputted from COMFORT is summarized in Figure 3.7. Only a few missions are repeated beyond the minimum requirements, e.g. PT2 (Basic Fighter Maneuvers) and TM4 (Close Air Support). These missions have relatively low costs in our model, and at the

---

[2] Note that the scheduler in TREFF is based on mission requirements, so we have to make sure the competency requirements are fulfilled as well. Also, the competencies with short retention intervals are sensitive to exact timings, but the current scheduler does not optimize to spread out the training optimally. Nevertheless, we do expect that a certain excess of sorties will be necessary. In a real system, one should adapt dynamically to the needs of each pilot, and would therefore not have to specify such a redundancy beforehand.

same time they cover many competencies. There are only a few excess repetitions of the complexity factors. This is to be expected since they are optional to most missions.

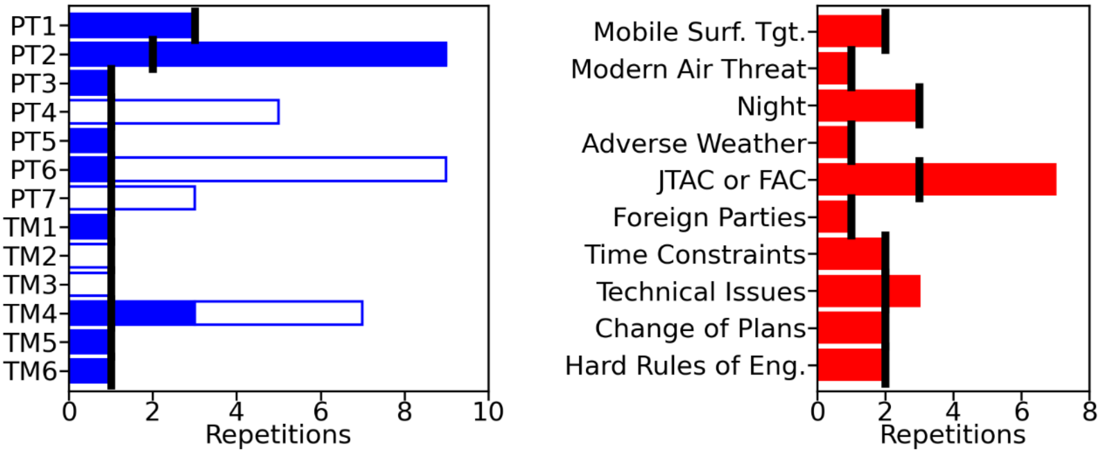


Figure 3.7 Number of repetitions of each mission (left) and complexity factor (right) in the list of training sorties. The full blue bars represent live training in aircraft, and the blue outlines represent simulator training. The black bars indicate the minimum requirements derived from the retention intervals.

The training program is broken down further in Figure 3.8, showing how the different missions add up to fulfill the competency requirements. For most of the (non-optional) competencies, the pilots get significantly more repetitions than the minimum requirements. The exceptions are the competencies with the shortest retention intervals (Managing Radar, Managing Electronic Support Measures, Handling Infrared Sensors, and Handling Targeting System Air-to-Air/Air-to-Surface). As for the complexity factors, there is no excess training of the optional competencies.

After we find the optimal set of sorties, COMFORT will reorder them in order to distribute the training for each competency as evenly as possible. This ensures that the pilots train each competency as close to the end of their retention intervals as possible. However, since each mission involves a different combination of competencies, it is in most cases not possible to get a perfect distribution for all competencies at the same time. This is accounted for by the redundancy factor C that we introduced above. The ordering of missions is implemented as a separate COP, where the cost for each competency is proportional to the longest sequence of items in the training program not containing the competency, and the required number of repetitions.

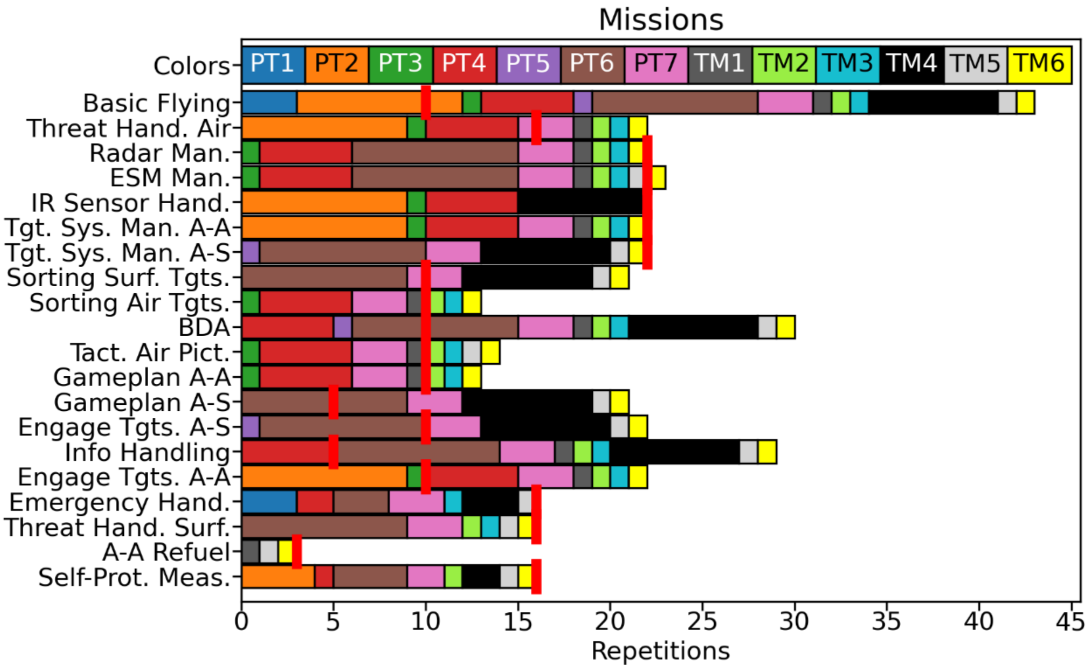


Figure 3.8 Breakdown of training program, showing which missions contribute to each competency. The colors indicate which mission contributes. The red bars denote the minimum requirements (including the redundancy C). Note that not all sorties with the same mission give the same contribution: they may differ in the optional competencies.

3.3 Training system specification

In addition to the training program, the TREFF input includes a number of properties of the squadron we are simulating training for and the availability of training resources. These are specified along with the training program in an Excel spreadsheet; see [8].

In the test case, we consider one squadron of 18 pilots. This squadron has one available aircraft slot per day. The number of aircraft available is given by a truncated Gaussian distribution with a mean of 8 aircraft available and a maximum of 10. Aircraft training is also affected by a stochastic factor representing bad weather. There are two daily training slots in the simulator for the squadron. There are 6 simulators available, and each simulator has a 98% chance of working for each sortie.

We also specify how many days the pilots have available for training. By default, the pilots work 5-day weeks, but TREFF can be configured to reserve time for other activities. We use this option to reserve one day a week for common briefings, administrative issues, and physical training. We include three different types of absence: four-week holiday, single-week leave, and one-day absence. They represent holidays, days with sick leave, non-flying training courses, and more. The absence is modelled using stochastic variables. Each pilot has one four-week holiday within the period from May to August. At most two pilots begin their holiday each week. The rest of the year, each pilot has a 15% chance of going on a one-week leave, this is drawn at the start of each week. Each pilot can have up to 4 such leaves each year. Finally, at any given weekday, the pilots have a 5% chance of one-day absence.

3.4 Results from Simulation Experiments

We configured TREFF according to the parameters given in section 3.3, and made a full-year training program consisting of two repetitions of the half-year program described in section 3.2. We simulate the full-year training program 100 times consecutively (we simulate 100 years), without resetting the proficiencies in between. The pilots have to wait for everyone to complete the half-year program before they can start the next repetition, both at the start of the year and mid-year. This remedy is necessary because the scheduler used in the test case does not distribute the training sufficiently well across the year otherwise.

The average number of days needed for everyone to complete the annual training program was 355 days; see Figure 3.9 (a). The length varies significantly, but only three runs take longer than a full year. The average pilot spends around 70 days on live training, 50 days on simulator training, and 85 days on other working activities; see Figure 3.9 (b).

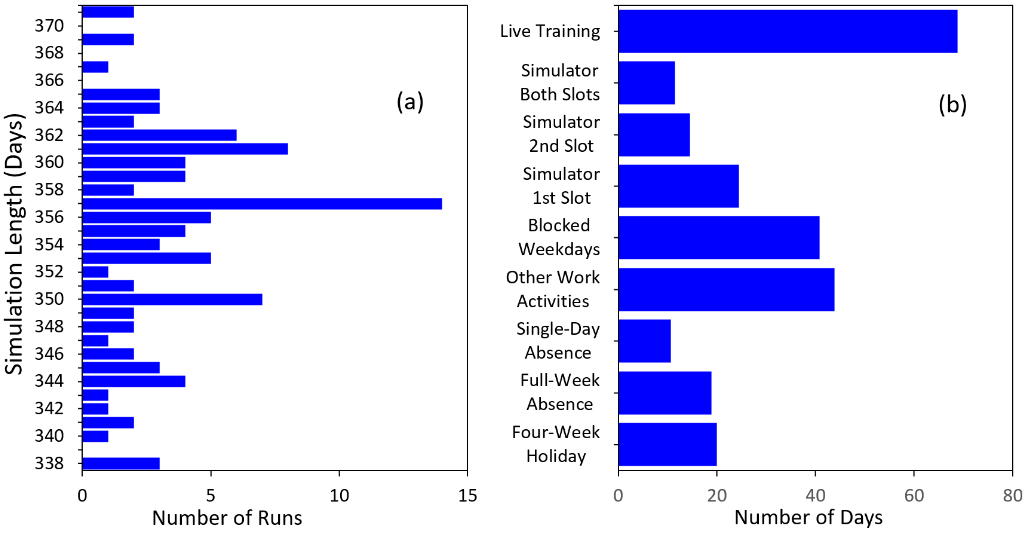


Figure 3.9 Simulated time in days for each repetition of the annual training program (a), and number of days used for each pilot activity (b). Red-air sorties are included in the live training category. The activity breakdown is averaged over all runs and all pilots.

Figure 3.10 shows how many times the pilots repeat each mission (as part of a blue formation). The pilots train in the ideal maximum formation sizes more often in the simulators than live. This is because the scheduler in TREFF prioritizes simulator training and live training differently, and there are two time slots each day for the simulators compared to one for the aircraft. The pilots get more repetitions than needed for some of the missions in the simulator. They also get more valid training in the simulators than in the aircraft, even though they spend fewer days for simulator training, since they can do two simulator sorties in a day and they don't have to fly red-air in the simulators.

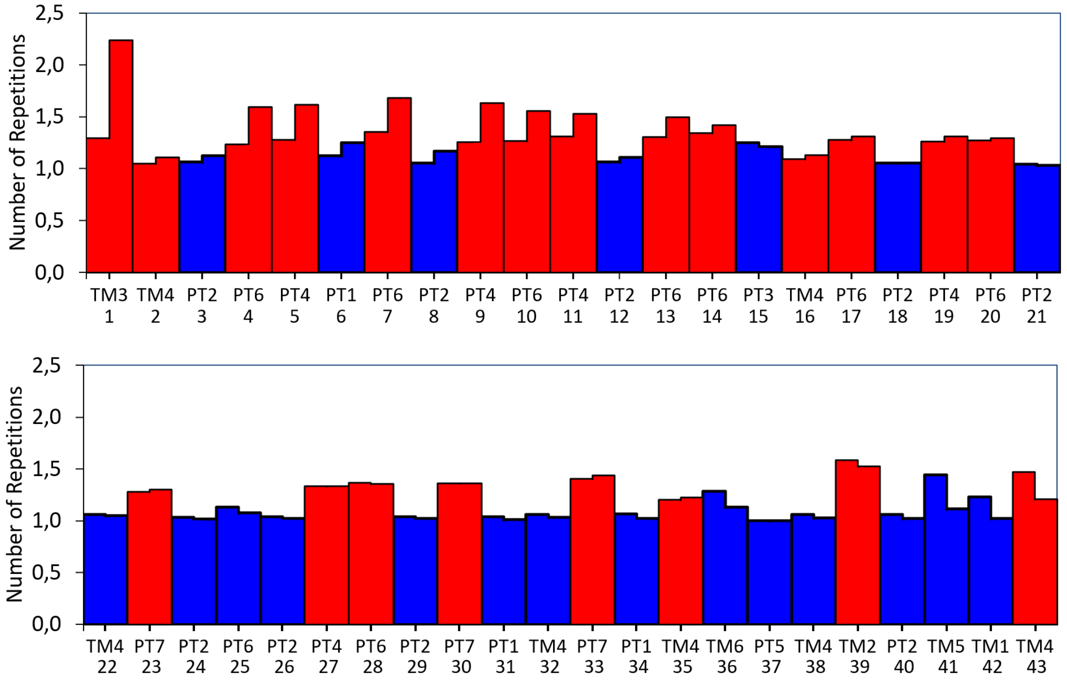


Figure 3.10 Average repetition numbers for each mission, split in two halves. The red and blue bars indicate simulator use and live training, respectively. The left half of each outline represents the first repetition of the training program, and the right half the second. Note that each repetition of the mission is here given an independent number, e.g. PT1 is trained as mission 6, 31 and 34.

TREFF calculates the proficiencies and generates plots during the simulation, giving immediate feedback about the training status of the pilots. These calculations include stochastic variations in the retention intervals and training outcome. To analyze the overall level of proficiency of the pilots, we calculate the average number of days per year where the proficiencies fall below the limit of 0.9; see Figure 3.11. The results show that for any competency, most of the pilots are proficient most of the time. Summing up the days with deficiencies in proficiency for each pilot, the total amounts to more than a year. This means that the pilots may in principle never be sufficiently proficient in all competencies, if they were only lacking one proficiency at a time.

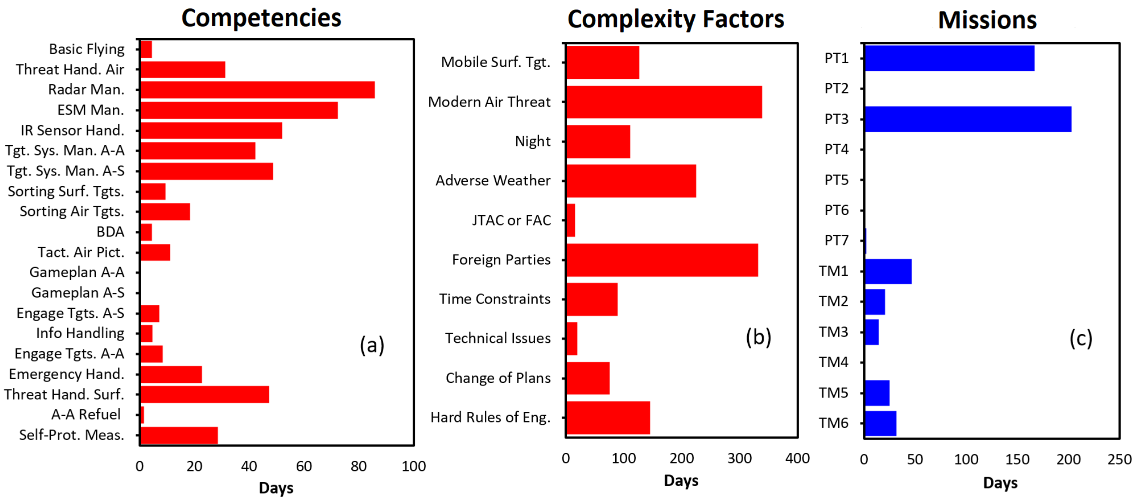


Figure 3.11 Average number of days per year where each pilot has proficiency lower than the requirement of 0.9, for each competency (a), complexity factor (b), and mission (c).

The proficiencies that fall below the requirements most often are linked to the competencies where the pilots get the least excess training; compare Figure 3.8 and Figure 3.11. For two of the missions, as well as six of the complexity factors, the proficiencies fall below 0.9 for more than 100 days a year. However, note that over the full year, the number of repetitions trained are equal to or higher than the repetition requirements. None of the proficiencies related to competencies fall below the requirement that much of the time. We prioritize the evenness of the proficiency profiles associated with missions and complexity factors lower than for the competencies, so they were not taken into account in the shuffling. The training program could be modified to include more training for missions and complexity factors, but this might lead to less efficient training of the competencies.

To get more insight into the drops in proficiency, we calculate the lowest proficiency among all the pilots for each competency. Figure 3.12 shows the behavior of these minimum proficiencies over a one-year period. As we see from the figure, rather than random drops in proficiency spread over the year, the pilots lose proficiency in many competencies at the same time in certain parts of the year. But for a significant portion of the year, all pilots are sufficiently proficient in all competencies.

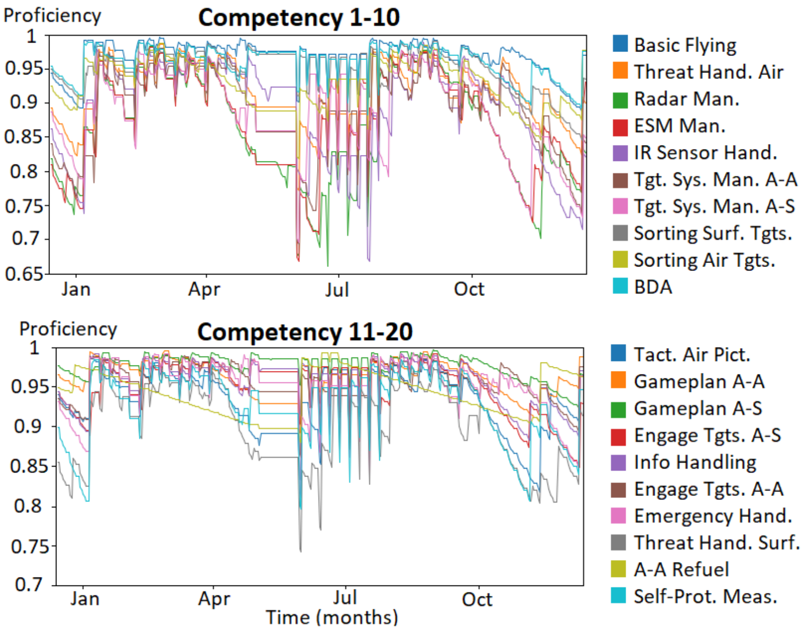


Figure 3.12 Minimum proficiencies among all the pilots as a function of simulation time, for an arbitrarily chosen year. The plots were directly generated in TREFF, which is implemented in AnyLogic [12]. Please note the different scales on the vertical axes.

The pilots start new repetitions of the half-year training program around January and July. Around these dates, several of the proficiencies fall significantly below 0.9; see Figure 3.12. This is because some pilots have to wait for others to complete the training program. All the pilots have to complete the entire training program before anyone can start the next repetition. All the pilots go on a 4-week holiday at some point between June and September. Their proficiencies are not updated when a pilot is on leave, but they are corrected for the entire period when the pilot resumes training. Therefore, we see some flat sections and large jumps in the minimum proficiencies around summer.

To estimate the impact of the individual variations in the retention intervals, and the variable training outcome, we ran a baseline simulation where all the pilots and training sessions where equal. In the baseline simulation, all the retention intervals are equal to the mean values given in Table 3.2, and the pilots always regain full proficiency after training. In the full retention model with individual variations, some pilots will inevitably require more training than the average. Therefore, we expect that the minimum proficiency among the pilots will improve when we remove the variations. Furthermore, the training outcome in the baseline retention model is always optimal, whereas in the full model pilots may need several repetitions to become fully proficient.

Figure 3.13 shows the minimum proficiencies for the baseline model, and is analogous to Figure 3.12 for the full model. The drops in proficiency are significantly less severe in the baseline model, particularly around New Year when the pilots start from the beginning of the training program. This indicates that the stochastic variations lead to larger competency deficiencies since we do not adapt the training accordingly. However, the training outcome in the baseline model has a higher average than in the full model, in addition to having less variations, which could also reduce the deficiencies. In any case, a more dynamic scheduler should be able to reduce the negative impact of both the variable training outcome and the individual variations by continuously adapting to the needs of the individual pilots.

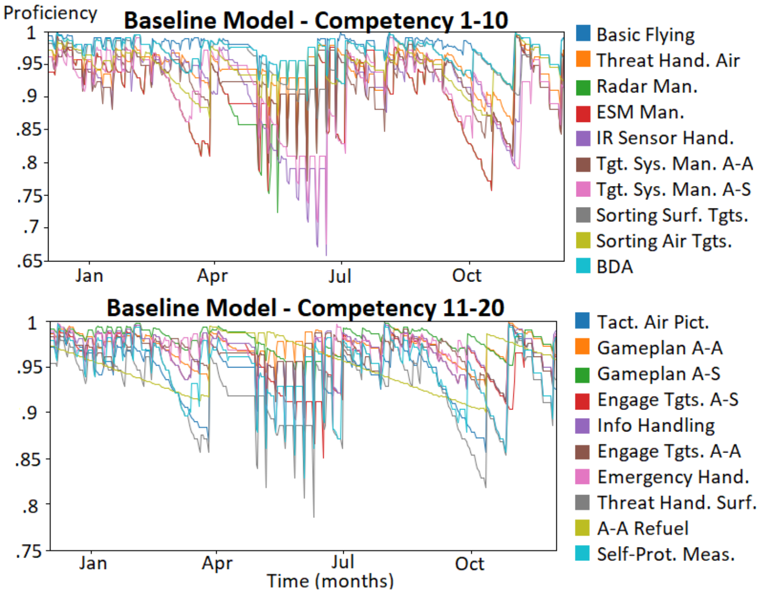


Figure 3.13 Baseline simulation without stochastic variations of retention intervals and training outcome. The graphs show the minimum proficiencies among all the pilots as a function of simulation time, for an arbitrarily chosen year. The plots were directly generated in TREFF, which is implemented in AnyLogic [12]. Please note the different scales on the vertical axes.

3.5 Test-Case Discussion

The scheduling problem in a competency-based pilot training system such as in the test case is challenging, not only because of the great flexibility, but also due to limited training resources and high co-dependency among the pilots. Our approach for the test case was to construct a static currency-based training program and use the existing scheduler from TREFF simulations. The COMFORT tool ensured that the training program gave sufficient training to maintain proficiency.

Our one-year training program takes on average around 355 days for everyone to complete, and therefore seems to be rather well matched with the available resources. Note that as we reserve one day a week for common briefings, administrative issues, and physical training, there is some margin that can be exploited. The TREFF scheduler attempts to complete all the training as fast as possible, and gives limited control over at what time and in which order the missions are completed. By implementing a scheduler that uses the competency requirements directly, one could improve the distribution of the training over the year. This approach should be explored for an analysis of a real system. However, the test-case results should still give a good indication of whether our requirements are feasible, and may help us identify some of the aspects that should be incorporated in such a scheduler.

In the test case, we include mostly high-level competencies, and only a few of the more specific constituent competencies. By breaking down the high-level competencies into constituent competencies, the competency-based training requirements can be made more complete and realistic. In doing so, we would likely find that the pilots need to practice many different missions and scenarios, whereas in the test case only a few missions are needed. Further analysis, discussions with operational pilots, and actual experiments are crucial in order to determine the most important training requirements.

One of the main purposes of the simulation is to assess whether it is feasible to maintain sufficient proficiency across the squadron over time with the given training resources. We added additional output channels allowing us to track and analyze data related to CBT. The proficiencies tracked during the simulation include individual variations modeled using stochastic variables. The training program is the same for all pilots. Thus, our simulation may be used to illustrate drops in proficiency due to individual variations that could be alleviated by individual adaptations.

In the test case, we evaluate the overall proficiency of the squadron based on the amount of time that the proficiencies for the competencies are less than 0.9. This may not reflect the actual needs and the policy of an air force. For instance, the requirements may be specified in terms of how many pilots should be combat ready for a specific mission, or how many can be trained to combat-ready status within a given time frame. Such operational requirements will be central to a future competency-based scheduler. Requiring that no proficiencies ever drop below 0.9 at all is likely too strict, and would require a disproportionate amount of training resources to maintain.

The pilots fall below the 0.9 proficiency requirements mostly around the time that the pilots finish the half-year training program. A scheduler that uses the competency requirements directly, as discussed above, would not necessarily give any such low-activity periods. The 4-week holidays during summer also lead to significant loss of proficiency, which would occur even with perfect scheduling and resource availability. For real analyses, it may be convenient to treat such periods of increased absence separately, and to allow the scheduler to plan the absence rather than model it as a stochastic process. An approach to distinguish between planned and unplanned absence is used in TREFF2 [9].

Co-dependency among pilots, unexpected events and limited resources will be challenging for optimization of combat aircraft training, even with support of an ideal scheduler. Therefore, one cannot determine whether a training system is able to fulfill a set of training requirements by looking at the requirements alone. A simulation such as TREFF allows us to take scheduling issues as well as the training resource availability into account, to design a system that is coherent with the training objectives and ambition. The TREFF simulation also provides a natural way to illustrate the workings of the test-case training system, and enable us to explore a large variety of alternative training systems.

For the basic test case, we used a linear skill decay model. More sophisticated models should be used for an evaluation of a real system, but a simple model is useful to understand the potential of our approach.

4 System concept and roadmap for the full PBT-approach

PBT is a data-intensive approach to support competency-based training. Competency-based training is a top-down approach to generate syllabi that are suitable to train for jobs with complex competency structures. PBT adds bottom-up fine-tuning to competency-based training. PBT is not just about optimizing a schedule, it may also provide the means to optimize and personalize the training content, at least when using a range of training devices. PBT, in a perfect world, would start with an ideal initial training (including Mission Qualification Training - MQT), which means there has been abundance of whole-task missions in a variety of complexities and situations. The result will be that each pilot is fully combat ready, has achieved higher-than-combat ready personal skill levels, and can apply the skills in a variety of contexts. Both retention and transfer of skills are as good as it gets for inexperienced pilots. While this ideal was nearly impossible to achieve a few years ago, with modern simulators it is within reach. Such training may be more expensive than ‘old school’ Initial Qualification Training/MQT, but the costs will pay back during the so much longer continuation training phase. This is because pilots only will be retrained when needed, and a considerable part of training may be in simulators. Note that it is recognized that live flying remains of vital importance in any phase of the pilot’s career. Although the inexperienced pilot will continue to grow skills while practice continues, the base level is already such that training may focus more on refreshing skills than on improving them.

In previous chapters, we outlined a PBT vision for personalized refresher training and explored how the simulation tool TREFF might provide analyses to enable this. The experiments did not yet implement every detail of the vision, and it may not even be advised to try and implement an idealized training concept from the start. There are ways to partially implement PBT and follow a safe and stepwise CD&E approach. In this chapter we will provide a global system concept for the end state. The chapter will end with some initial options for a roadmap with partial implementation.

4.1 A PBT system for Continuation Training

The complexity of PBT requires the training system to be extended by a set of support functions. These should enable analysis and presentation of current and expected training needs, supported by automation and advanced modelling. The aim is that advised training schedules for a given period (ranging from weeks to a full year) will enable an optimum over the various personal training needs and team/collective skills requirements, given constraints such as availability of resources (human, training media), expected leaves and special events. For example, large-scale exercises need to be planned for well in advance, and should therefore be considered as given. They will, however, constitute training opportunities for individuals.

Key components of the PBT system will be the following: a quantitative personal retention model, an advanced scheduling tool, a dashboard to present the appropriate selection and detail of results to a variety of users, and a database for all relevant information to calculate the optimal training and timing for each competency, complexity factor, and task/mission.

Quantitative Personal Retention Model

This model will calculate the expected retention curve for each specific task, complexity, and/or competency/skill-based on 1) the track record of personal performances and 2) generic retention influencing factors. For each element, a desired minimum of retention can be determined. The predicted retention will have some level of uncertainty.

What counts as a minimum, results in some cases from legislation, but more often from national or operational policy, and may further be guided by cost-benefit analysis. There may be good reasons to allow pilots to fade their skills to for example a 50% proficiency level and only when needed to fully refresh to 100%. This will require a longer period of refresher training compared to a training regime in which pilots would never fall below 90% proficiency. The total costs for a variety of training regimes can be compared.

Scheduling tool

The retention model generates a considerable number of retention curves and minimum retention intervals per person. It will be a challenge to ensure all these minima are met (or minimally exceeded) for an individual pilot, let alone for the full squadron, wing, or air force. Optimization algorithms may produce optimal retention intervals taking care of individual needs, but optimized over the full group while considering expected options/constraints on resources (aircraft, airspace, simulators, and support personnel), work shifts, legislation, absence, other squadron tasks, team demands for the various missions, etc.

The tool may have several functions. It may be used for resource planning in combination with a simulation to determine the feasibility to completing training for all individuals over the year given desired minimum readiness / proficiency levels. It may also generate an actual schedule for a given next period of time. The time scale therefore needs to reflect the planning horizon and purpose of the scheduling. In addition, the tool should be used to handle the set of deviations that always occur, such as illness, aircraft or simulator failure and bad flight weather.

At the beginning of a new planning horizon, the scheduling tool should take into account what was actually being trained by whom in the previous period.

Dashboards & interfaces

PBT is meant to optimize training and enable informed decisions throughout the year and on every level in the training system, from the individual pilot to the full air force. This implies that a PBT system will have several users. Individual pilots may want to monitor their weak and strong points, check the focus of coming sorties, or reflect and respond to certain system generated proficiency scores. This will require the PBT system to provide for a dashboard that is tailored to the needs of the pilot. This may be a completely different dashboard than needed by e.g. a wing commander. They may not have access to detailed individual pilot data, but only their current or predicted readiness levels as well as squadron wide readiness levels. The personnel responsible for scheduling training missions obviously need a dashboard for monitoring and adjusting a training schedule.

Note: It may be tempting to use the PBT system for additional motivation and fun in the form of gamification. A public leaderboard may be used to reveal who scores highest on certain competencies or on a mix of such competencies. Such practice entails certain risks. It may generate unwanted ‘playing the game’ behavior leading to unsafe flying or risky tactical situations. Also, as no one want to be at the bottom of the list, the lowest performers may seek to enhance their scores in undesirable ways. If a leaderboard is used, it is advised only to enlist one to three top performers. Also, a leaderboard may enhance competitiveness, but may undermine group cohesion and the tendency to support each other in the process of retaining and improving proficiency.

Database

PBT requires a very extensive database for determining retention intervals (performance and cognitive load metrics, personal factors, training (situational) factors), modelling results (predicted retention, retention intervals, and advised schedules), and actual/predicted constraints for scheduling (availability data, predicted sick leave, vacation, flight time limitation legislation, weather conditions).

Foreseen user requirements

A PBT system should have the above components with outlined functionalities. In addition, we assume several ‘user requirements’:

1. It should be possible for the officer responsible for scheduling to overrule a recommended training event, a recommended training device, and/or a recommended training slot.

2. The scheduling tool should be able to reflect the necessary planning horizon

3. The system (here the retention model) should respond automatically to actual training events (pilot X has performed sufficient on competencies A, B, C and insufficient on competency D; given complexity factors F and G on training device M)

4. The system can perform with missing data (e.g. missing assessments)

5. The scheduling system is able to make proposals to handle deviations from plan (e.g. missing aircrafts, pilot illness)

6. The predicted skill decay and its resulting retention period may be adjusted manually (by e.g., supervisors / instructor pilots / flight commanders) in case they consider the computational personal retention model flawed.

Obviously, the user requirements will need to be validated and completed in consultation with the operational community.

4.2 From the squadron perspective: how to start evolving towards PBT

The idealized PBT system outlined above may be positioned on a 10-year horizon that we may strive to reach, but along the journey we should try to implement what is feasible, safe, and acceptable. After all, training based on performances, rather than missions with a given frequency, entails a fundamental shift in the design and scheduling of training. It is likely that several steps will have to be made before reaching a system that is truly based on performances. In this section, we outline our first notions of a roadmap towards PBT in a feasible way, trying to take a squadrons perspective instead of an academic one. In the Netherlands, partial implementation of these steps has been achieved in the Defence Helicopter Command. A different, more rigorous implementation of PBT can be found in the United Stated [13]. Lessons learned of such approaches are not yet publicly available. The first step would be the creation of a full competency profile for combat aircraft pilots. The competency profile needs to be specific and reflect the requirements from national authorities. We recommend including complexity factors from the start, to account for the different conditions the competency must be applied under. The competency profile should cover the distinction between wingmen and flight leads (two ship, four ship, and larger formations).

The next step would be to estimate average retention intervals for competencies. It is known from practice that experienced pilots have significantly different retention profiles from unexperienced pilots in their training needs. It is therefore advisable to use different retention intervals for different experience groups.

The competency profile and associated complexity factors must then be mapped to retention intervals. The proficiency within a given competency is retrained in the context of a regular training program built up from (training) missions (part-task or full mission), to be trained in the aircraft or in a simulator. While implementing the training program, proficiency on these competencies may be scored, not from a formal grading perspective, but to start collecting data for later performance modelling.

A mapping between competencies, missions and complexity factors is therefore relevant. One option is to use surveys, e.g., with self-ratings, to validate such a mapping and the associated retention intervals. For conducting such surveys, it may be easiest to start by focusing on only a subset of competencies and complexity factors. This should provide estimates for how much competency is increased per mission carried out. It may also be easiest to start with simpler missions, and to use simulator missions to better control all aspects of the training, including complexity factors. The surveys can then be expanded until results start to show clear patterns. At later stages, self-ratings should be complemented by instructor ratings or more objective metrics of performance, e.g. data from a simulator. If the competency profile is hierarchical, the proficiency of high-level competencies should depend on the proficiency of the underlying constituent competencies.

Before starting to train according to competency-based requirements, it is important to clearly define updated overall training requirements. Are competency-based requirements to come in addition to the currency-based training requirements (flight hours, missions, specific events), or are they to replace some of those requirements? It is likely that these requirements will be developed as more experience with CBT is obtained, and a possible first step is to keep the existing requirements as hard requirements and add soft requirements on competencies. The CBT-requirements must describe the desired minimum proficiency profile, such as: “all pilots should have proficiency above x % for all competencies and complexity factors. In addition, 50 % of a prioritized set of competencies should have a proficiency level above y % “. They should also include more quantitative requirements like national requirements for the number of sorties.

Scheduling of training in order to minimize the competency gap of all trainees with the least number of total training sessions is complicated, and it is likely that a scheduling tool should be used, in order to be able to best match the trainees with a specific mission (including specification of complexity factors and inclusion of events enabling the optional competencies).

With support from a scheduling tool, it is possible to start exploring the feasibility of the requirements by means of simulation experiments. A simulation tool like TREFF2 can be used for this. This simulation tool should include the same scheduling tool described above.

Once the requirements for training are well defined, it is possible to experiment by adjusting the training based on CBT. A natural first step is to begin with only a limited group (e.g. one squadron) for a limited time, and only consider training in the simulator. The surveying system for evaluating proficiency should be maintained. The experimentation should be planned using feasibility simulation experiments. During actual experimentation, the scheduling tool should be used to advice in scheduling of training.

While experience builds up with CBT, predictive algorithms may be developed and tested on actual performance. Using predictive algorithms may be validated first in simulator sessions on part tasks. This may apply to a group of pilots or to individual pilots if sufficient data is available.

As experience with CBT and PBT grows, performance-based scheduling of training missions and full missions in the simulators may follow, and in a later stage, the smaller live missions as well. There will be external factors out of control for the scheduling tool, such as larger national and international exercises. The CBT concept may still guide the focus of training for individuals participating in such activities, and the scheduling tool used to determine which pilots should participate when to which exercise.

In the final stage, data may be gathered in increasingly more automated way and performance predictions and optimized scheduling may also be performed automatically.

5 Summary and conclusion

This work considers optimization of performance-based continuation training. A simple test-case is used to highlight opportunities and challenges for realizing a PBT training system for combat aircraft pilots, as a first step towards such an approach. The work is based on a joint approach, combining research from NLR on designing PBT training programs, with research from FFI on feasibility of training programs taking scheduling challenges and resource constraints info account. The test case used is based on example data, and the emphasis is on the method.

The following research questions were formulated for optimization of performance-based continuation training:

A. How can competency profiles be used to create a training program?

B. How will the flexibility of personalized training affect the feasibility of the scheduling?

C. How can competency profiles be used to create a scheduling tool?

Addressing these in full requires further work, supported by a PBT-specific scheduler based on operational insights and priorities. As a first step, we have investigated the following research questions:

I. How can the NLR approach to CBT be used to provide input for TREFF?

II. What are the possible challenges with the flexibility of personalized training, and what are the relevant metrics for a scheduling tool?

Question I) is a first step towards investigating how competency profiles can be used to create a training program (A). Training programs in the format required by TREFF are mission-oriented and focused on the resources needed (aircraft or simulator and pilots). Co-dependency among pilots and availability of resources is at the center of investigating the feasibility of the conducted training. From this perspective, it is not essential exactly which competencies are involved in a given mission or under which complexity factors the training is conducted. The underlying assumption is that given a sufficient number of training sorties for each mission, the flight-lead will be able to plan the exact content of the training so that the trainees fulfill their training needs. For competency-based training, the design of the exact content to best fit the needs of each trainee is at the center.

The starting point for creating of a TREFF training program from NLRs CBT approach is a set of missions, competencies and complexity factors with associated retention intervals. A mapping must then be conducted to relate the missions to the competencies and complexity factors. There is a large number of combinations of activities leading to the same competency levels. We therefore apply an optimization tool – COMFORT, to find the best suitable training program. The constraints and the target function to be optimized can be adapted so that the resulting training program best fits the operational priorities and the requirements for combat readiness. Note that the training program obtained in this manner is much more detailed than the CBT requirements from which it was obtained, leaving less choices up to the flight lead.

We find that this method is well able to define a training program that can provide the necessary input for an analysis using TREFF. However, this should be accompanied by operational insights regarding priorities of proficiency levels for the different competencies, skills and complexities. We also identified some possible refinements of the CBT input to more clearly define a training program:

1) Different levels of competencies should be clearly distinguished. In the current experiment, we have used mostly high-level competencies. However, more specific competencies, which may be applied less often and tend to be more vulnerable to skill or knowledge decay, should be identified and used for scheduling purposes. Using more detailed competencies and more realistic skill decay data will obviously make a project classified.

2) The estimated training outcome when conducting a mission should be evaluated, in particular how it is affected by the number of competencies trained and under which complexity factors. For example, a mission focusing on the competency of applying radar may also involve training on basic flying, but the training outcome for basic flying may be different than if that was the only competency being trained. Note from our mapping of missions and competencies that some missions include a large number of competencies.

3) The training requirements should focus on a limited set of competencies. Once it is established (or can be assumed safely based on current experience) that certain competencies are trained well before they decay, no specific measurements need to be taken and efforts to monitor performance may focus on a select set of more vulnerable competencies.

4) In our simplified case, we considered the pilots as one homogeneous group. For a true implementation, one should also consider the differences in experience-level and how this affects the retention intervals. Additionally, the competencies/complexity factors should reflect the differences in training depending on the role in the formation.

Question II (What are the possible challenges with the flexibility of personalized training, and what are the relevant metrics for a scheduler) is a first step towards defining a scheduler based on competency profiles, and to identifying how the flexibility of personalized training may affect the feasibility of the training.

To consider the effect of personalized training, we added a stochastic training outcome to each training sortie, and added a stochastic variation to the retention intervals. As the scheduling of the training is not based on competencies directly, the results from the simulation cannot immediately be translated to findings in this regard. However, the results clearly show the variations in proficiencies. There are large individual differences, especially for competencies with short retention and in periods just after vacations or other kinds of absence. Comparing simulations with and without variations in personal retention, preliminary results indicate that stochastic variations lead to larger competency deficiencies for certain pilots. This is not unexpected since the training is not adapted accordingly. Including measures on the competency profiles and complexity factors will provide a more realistic view of the training level, and provide an opportunity to better adapt the training.

Variations in competency profiles across the squadron will on one hand lead to more difficult scheduling as different missions will best fit the needs of the different pilots. On the other hand, compared to currency-based training, there will be an increased flexibility and many missions may contribute to increasing low levels in competency. Increased insight into 2) will be important to evaluate the effect of this flexibility.

A scheduling tool that uses competencies will be a necessary next step for optimized PBT. Depending on operational priorities, this scheduling tool could be based on a mixture of competency optimization along with specific sortie requirements. As outlined in the roadmap of chapter 4, the time horizon should be adapted to the time horizons of the squadron. The scheduling tool should balance individual pilot needs with squadron wide requirements. The scheduling tool of [9] could be a good starting point.

In the test-case considered in this report, the complexity factors are separated into inherent and optional. Some of these optional complexity factors, such as adverse weather, cannot be planned for when training in the aircraft. The scheduler should omit these when planning for live-aircraft sorties. Experience gained during live sorties may still be considered for scheduling simulator training. Optimal scheduling, in which all pilots always train for each competency, complexity factor and mission exactly when reaching the lower threshold of proficiency, will not be possible in practice. Always keeping the proficiency above this level to have some margin is one option, however this will lead to a redundancy that may be too costly in reality. Operative priorities are again important to reach a sound balance, where an acceptable training level is achieved without requiring too many additional sorties. Iterations with simulations like presented in this report may be useful to reveal the redundancy associated with a given set of requirements, and vice versa: the potential lack in proficiency if redundancy is reduced.

Transitioning to the full PBT-approach will entail a fundamental change in training concept, and we recommend a stepwise process in our roadmap towards the full PBT-approach. This will enable the squadron to gain experience, develop performance measures and retention models, and validate the resulting competency profiles. Scheduling the training will be an increasingly complex task with more degrees of freedom as mission specification is expanded to specification of missions, competencies and complexity factors. We therefore recommend the use of a scheduling tool to guide the squadron scheduler. When working towards this roadmap, we see potential in using a simulation tool such as TREFF/TREFF2 to plan experiments and next steps.

Appendix

A. Abbreviations in the test case

Complexity Factor Abbreviation	Complexity Factor Name / On-State	Off-State
Mobile Surf. Tgt.	Mobile Surface Target	Fixed Surface Target
Modern Air Threat	Modern Air Threat	Old Air Threat
Adverse Weather	Adverse Weather	Normal Weather
JTAC or FAC	JTAC or FAC	No JTAC and No FAC
Foreign Parties	Foreign Parties	No Foreign Parties
Time Constraints	Time Constraints	Relaxed Time Constraints
Technical Issues	Technical Issues	No Technical Issues
Change of Plans	Change of Plans	No Change of Plans
Hard Rules of Eng.	Hard Rules of Engagement	Easy Rules of Engagement

Competency Abbreviation	Competency Name
Basic Flying	Basic Flying Skills
Threat Hand. Air	Threat Handling Air
Radar Man.	Managing Radar
ESM Man.	Managing Electronic Support Measures
IR Sensor Hand.	Handling Infrared Sensors
Tgt. Sys. Man. A-A	Managing Targeting System Air-to-Air
Tgt. Sys. Man. A-S	Managing Targeting System Air-to-Surface
Sorting Surf. Tgts.	Sorting Surface Targets
Sorting Air Tgts.	Sorting Airborne Targets
BDA	Battle Damage Assessment
Tact. Air Pict.	Building Tactical Air Picture
Gameplan A-A	Gameplan Execution Air-to-Air
Gameplan A-S	Gameplan Execution Air-to-Surface
Engage Tgts. A-S	Engage Targets Air-to-Surface
Info Handling	Information Handling
Engage Tgts. A-A	Engage Targets Air-to-Air
Emergency Hand.	Emergency Handling
Threat Hand. Surf.	Threat Handling Surface
A-A Refuel	Air-to-Air Refueling
Self-Prot. Meas.	Use of Self-Protection Measures

References

[1] W. Arthur Jr. *et al.*, "Factors that influence skill decay and retention: A quantitative review and analysis," *Human Performance*, vol. 11, no. 1, pp. 57–101, 1998.

[2] J. I. Vlasblom *et al.*, "Competence retention in safety-critical professions: A systematic literature review," *Educational Research Review*, vol. 30, p. 100330, 2020.

[3] *Understanding and Reducing Skill Decay*, NATO Research Task Group HFM-292, final Report, in prep.

[4] E. Svensson *et al.*, *Skill Decay, Reacquisition Training, and Transfer Studies in the Swedish Air Force: A Retrospective Review*, 03 2013, pp. 258–281.

[5] J. van Merriënboer and P. Kirschner, *Ten Steps to Complex Learning: A Systematic Approach to Four-component Instructional Design (Third Edition)*. Routledge, 2017.

[6] C. M. Colegrove and G. M. Alliger, "Mission essential competencies: Defining combat mission readiness in a novel way," in *RTO-MP-99 – NATO symposium Air Mission Training Through Distributed Simulation (MTDS) – Achieving and Maintaining Readiness*, 2002.

[7] A. M. Law, *Simulation Modeling And Analysis (Fourth Edition)*. McGraw-Hill, 2007.

[8] G. Skogsrud and O. M. Mevassvik, "Simulation of a combat training program for Norwegian pilots," in *RTO meeting proceedings of the NATO Modeling and Simulation Group*, Bern, Switzerland, Oct 2011.

[9] G. K. Svendsen *et al.*, "Optimized pilot training for combat aircraft," in *Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2017*, no. 17199, Orlando, FL, Dec 2017.

[10] E. L. Fjærbu, "Optimization of training programs for competency-based training," Norwegian Defence Research Establishment (FFI), FFI-rapport 20/00002, 2020.

[11] Google, "Google OR-Tools," 2021. [Online]. Available: https://code.google.com/p/or-tools/

[12] The Anylogic Company, "Anylogic," 2021. [Online]. Available: https://-www.anylogic.com

[13] "A new us air force jet pilot training construct," *The Military Simulation And Training Magazine*, vol. 37.

About FFI

The Norwegian Defence Research Establishment (FFI) was founded 11th of April 1946. It is organised as an administrative agency subordinate to the Ministry of Defence.

FFI's mission

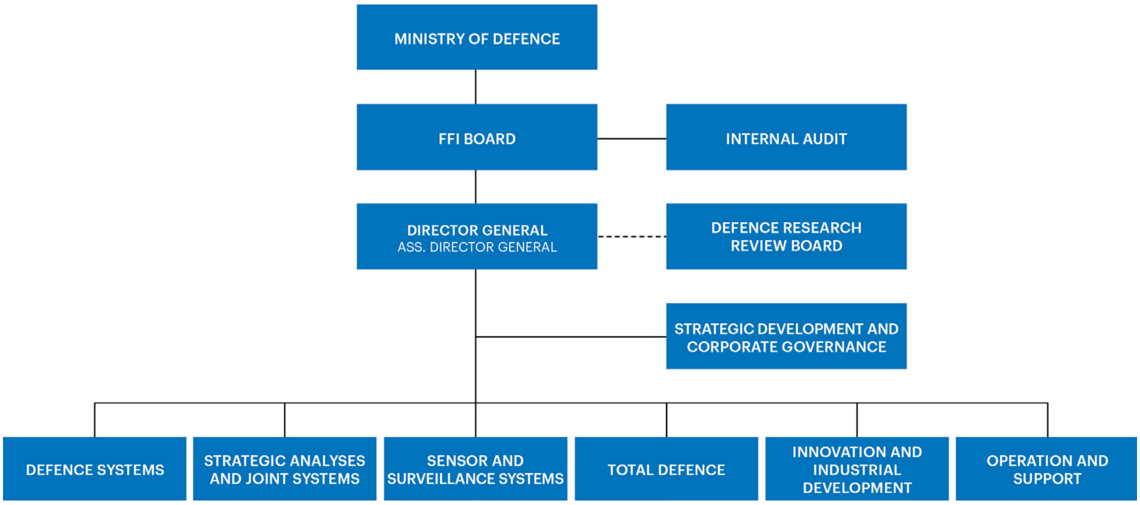
FFI is the prime institution responsible for defence related research in Norway. Its principal mission is to carry out research and development to meet the requirements of the Armed Forces. FFI has the role of chief adviser to the political and military leadership. In particular, the institute shall focus on aspects of the development in science and technology that can influence our security policy or defence planning.

FFI's vision

FFI turns knowledge and ideas into an efficient defence.

FFI's characteristics

Creative, daring, broad-minded and responsible.



Forsvarets forskningsinstitutt
Postboks 25
2027 Kjeller

Besøksadresse:
Instituttveien 20
2007 Kjeller

Telefon: 63 80 70 00
Telefaks: 63 80 71 15
Epost: post@ffi.no

Norwegian Defence Research Establishment (FFI)
P.O. Box 25
NO-2027 Kjeller

Office address:
Instituttveien 20
N-2007 Kjeller

Telephone: +47 63 80 70 00
Telefax: +47 63 80 71 15
Email: post@ffi.no