

Ludwig Gelot*, Zoran Todorovic

Evaluating joint operations in staff exercises: A novel method for assessing training objective fulfilment in Swedish-Finnish command-post-exercises

DOI 10.2478/jms-2025-0002

Received: August 21, 2024; Accepted: January 10, 2025

Abstract: Sweden and Finland have a long history of cooperation in military training, pooling resources to educate officers for NATO crisis response and United Nations (UN) peace operations. Key exercises like the Combat Joint Staff Exercise (CJSE) and VIKING (VK) are essential for training officers to handle complex conflict environments. Despite their success, the current evaluation framework for these exercises lacks specificity and reliability, leading to general and insufficiently detailed assessments. This article reviews the existing evaluation methods used in CJSE and VIKING, highlighting their strengths and weaknesses. It identifies issues, such as unclear objectives and insufficient criteria, which result in gaps and overlaps in evaluator reports. To address these problems, this article proposes a new evaluation framework based on Kirkpatrick's criteria, incorporating specific, measurable, assignable, realistic and time-related (SMART) training objectives (TOs) subdivided into specific training objectives (STOs) and microlearning objectives (MLOs). This new approach aims to provide a more detailed and aggregable evaluation, supporting timely decision-making by exercise directors and control centres.

Keywords: evaluation framework, SMART, learning objectives, performance metrics, Kirkpatrick's criteria, specific training objectives

1 Introduction

Sweden and Finland have a long history of cooperation in training and exercising their respective armed forces.

Faced with shared geostrategic threats, they have effectively pooled resources to educate their aspiring officers (Depledge 2020). Such distributed staff exercises include the Combat Joint Staff Exercise (CJSE) and VIKING which are designed to prepare the officers educated at the Swedish and Finnish Defence Universities to be deployed in NATO crisis response operations and UN peace operations.¹ These exercises are essential to ensure that officers are well-prepared to contribute effectively to complex and dynamic conflict environments (Enstad 2022; Roennfeldt 2022, p. 194).

CJSE and VK exercises have been successfully organised for over two decades, and final exercise reports (FER) consistently point to their usefulness in building staff competencies. Evaluators state that the exercise objectives (EO) are met to a satisfactory standard. However, while the EOs may be met, the system in place to evaluate the fulfilment of training objectives (TO) has not been able to provide a measurable, timely and reliable overview of the progress made by the training audience (TA) on a daily basis. The lack of a consolidated evaluation framework with specific objectives, timing and criteria has meant that evaluations tended to be parsimonious and overly general.

This article explores the evaluation framework used to assess trainee performance in CJSE and VK staff exercises. It reviews the strengths and limits of an evaluation process that relies on Hot Wash Ups, After Action Review (AAR), and active mentoring and training by observer, trainer, mentors (OTM). It describes how unclear evaluation objectives have led to gaps and overlaps in the reports of the evaluators. Insufficiently detailed evaluation criteria with missing benchmarks, timings, standards and thresholds led to difficulties in the standardisation of

*Corresponding author: Ludwig Gelot, Joint C5 Warfare Centre, Joint Signal Regiment, Enköping, Sweden, E-mail: l.gelot@outlook.com
Zoran Todorovic, Joint C5 Warfare Centre, Joint Signal Regiment, Enköping, Sweden

¹ It is organised annually but has been replaced at times by a learner alternative during the COVID pandemic.

measurements. Ultimately, this meant that observations have been insufficiently specific and difficult to aggregate in order to illustrate the overall progress made by the TA on a daily basis.

This article reconsiders the CJSE and VK evaluation framework and describes a standardised approach to assess the fulfilment of TOs. Following a review of Kirkpatrick's (1976) evaluation criteria, it describes concretely a more detailed TO evaluation process that facilitates reporting and measurement. Through the development of specific, measurable, assignable, realistic and time-related (SMART) TOs subdivided into specific training objectives (STO) and micro-learning objectives (MLO), it describes an aggregable evaluation framework suitable to support the timely decision-making process of the exercise director and exercise control centre (Doran 1981).

2 Objectives in CJSE and VK

The CJSE and VK Command-Post Exercises (CPX) are supported by shared political and academic interests. They aim to train officers in the joint functions that characterise high-intensity warfare in a NATO Crisis Response Operation (CRO) and a UN peacekeeping operation, respectively. Since 1999, Sweden has organised the world's largest UN peace operation exercises called VIKING (Gelot 2019). The latest edition, VIKING 22 (VK22), saw military, police and civilian personnel from 47 different nations train simultaneously across 8 sites in 5 different countries.² A total of 1,450 persons were in the primary and secondary TAs in 9 headquarters. The first CJSE was conducted in 2005, and it has been organised on a quasi-annual basis. The latest edition, CJSE21, saw close to 1,000 persons train a NATO CRO in a high-intensity environment.

The aims and objectives of CJSE and VK exercises are outlined in their respective Exercise Terms of Reference (EGToR) and Exercise Specifications (EXSPEC). These documents are used to subsequently develop TOs in the Exercise Plan (EXPLAN). Exercise and TOs are then evaluated, and results are reported in the FER published in the wake of the exercise.

EOs refer to the overarching goals or outcomes that a military exercise aims to achieve. These goals are often strategic or operational in nature and are aligned with the

broader mission and goals of the military organisation or government ordering the exercise. EOs are supposed to guide the planning and execution of the exercise. For example, CJSE21 had the following EOs:

- Enhanced understanding of planning and execution, including assessment, of a Combined and Joint CRO.
- Enhanced individual ability to participate and understand a staff's internal processes and procedures in accordance with Standard Operating Procedure (SOP).
- Enhanced individual ability to act as staff members, branch heads and commanders in an international staff environment during a Combined and Joint CRO.

VK22 had broader EOs that also covered experiments conducted during the exercise:

- Promote mutual understanding, confidence, cooperation and interoperability among all contributing and affected forces, organisations, offices and personnel – military as well as civilian (i.e. comprehensive approach).
- Understand and apply mission command and management, staff roles and functions, procedures and structures, as well as coordinated planning processes.
- Understand and apply current operational concepts reflecting present as well as future challenges in multinational and multidimensional peace operations.
- Create an environment that supports and facilitates development and experimentation of methods, operational concepts and technological enhancements for participating organisations and nations.

TOs are specific and measurable goals set for the individual units, personnel or teams participating in the exercise. These goals are designed to enhance the skills, knowledge and ability of the participants. For example, they may involve goals related to the implementation of SOPs and processes for joint functions within a NATO CRO. CJSE21 had the following TOs:

- Conduct current operations, including planning, execution and assessments, coordinated with relevant actors in accordance with valid SOP, Operational Plan/Joint Coordination Order/Fragmentary Order (OPLAN/JCO/FRAGO), relevant documents and Component Commands (CC).
- Conduct Mid-Term Planning, including planning, execution and assessments, coordinated with relevant actors in accordance with valid SOP, OPLAN/JCO/FRAGO, relevant documents and CC.
- Conduct Long-Term Planning, including planning, execution and assessments, coordinated with relevant

² Sweden, Finland, Brazil, Bulgaria and Qatar. Teams from Ukraine and Bosnia and Herzegovina participated in the planning phase but could not conduct the exercise in their respective countries.

actors in accordance with valid SOP, OPLAN/JCO/FRAGO, relevant documents and CC.³

EOs and TOs are the subject of a complex evaluation process that operates on two primary tiers. It relies on an evaluation function (EVAL) responsible for assessing the fulfilment of EOs by identifying strengths and weaknesses in both the planning and execution phases of the exercise. This includes the evaluation of the adequacy of the exercise in enabling the TA to meet their TOs.

The second tier of evaluation is covered by the OTMs. OTMs are Subject Matter Experts (SMEs) tasked to facilitate the TA learning process and to ensure that participants meet the TOs. OTMs report observations on the fulfilment of TOs. EVAL will rely on assessments from the OTM organisation and self-reports from the distributed sites regarding TA performance and fulfilment of TOs and does not directly participate in the evaluation of TA performance.

In this article, we focus on the evaluation framework used to assess the learning process of participants and the fulfilment of TOs. The other tiers covered by the EVAL function are not addressed since they focus on the adequacy of the exercise structure rather than actual learning. In the following section, we review the evaluation framework used in CJSE and VK and analyse how Kirkpatrick's Four-Level Model for training evaluation applies to these staff exercises. This will enable us to identify limits in the application of the current evaluation process, thereby paving the way to the outline of a revised method.

3 Evaluation framework

CJSE and VK can be defined as collective training aiming to train adult learners in job skills to improve professional performance. In this context, training evaluation can be defined as 'a system for measuring whether trainees have achieved learning outcomes. It is concerned with issues of measurement and design, the accomplishment of learning objectives, and the attainment of requisite knowledge and skills' (Goldstein 1986, p. 312). To enable this evaluation process, a framework must be chosen due to the sheer multitude of learning outcomes that could be considered part of training evaluation (Kraiger et al. 1993, p. 323).

Simpson and Oser (2003, p. 27) define an evaluation framework as 'a set of evaluation events with prescribed

objectives, timing, evaluation methods, and evaluation criteria'. CJSE and VK are such complex learning events that evaluators need to focus their attention on very specific areas. Indeed, learning outcomes for evaluation could possibly cover anything from cognitive outcomes such as verbal knowledge to affective outcomes such as attitudinal change or skill-based outcomes (Simpson and Oser 2003, p. 27). Moreover, evaluation could occur during different timings and phases (Simpson and Oser 2003, p. 33). The evaluation criteria can also vary to cover TA reaction, learning, behaviour or results (Kirkpatrick 1976).

Various evaluation frameworks exist to assess military units. NATO's Combat Readiness Evaluation (CREVAL) is a comprehensive assessment process used to evaluate and ensure the combat readiness of military units within the NATO alliance. CREVAL includes thorough evaluations and exercises testing a unit's readiness, covering operational capabilities, equipment, personnel proficiency, command structure, interoperability with allies and adherence to NATO standards. In Sweden, the Swedish Armed Forces use Military Analysis Method for Reliable Tactical Assessments (MARTA) as a strategic evaluation approach. They involve on-site observers assessing military units based on six key capabilities. This method helps in meticulous documentation, pinpointing areas for focused training in subsequent exercises.

The main issue with the above-mentioned models is that they are applied to standing units and become irrelevant in the context of staff exercises for temporarily composed staff such as CJSE or VK. Furthermore, they are not designed to evaluate the capacity of a headquarter to implement the type of joint processes necessary in a CRO or a peace operation. For example, MARTA will check the existence and quality of standing orders, but not the capacity of a unit to work in an integrated manner as part of joint functions such as joint fires or logistics.

CJSE and VK employ an alternative evaluation method with a team of OTMs, typically from the teaching staff of the Swedish and Finnish Defence Universities. OTMs are typically seasoned military personnel, instructors or SMEs who possess a deep understanding of the TOs, exercise scenarios and the specific processes being tested. OTM responsibilities include:

1. **Real-time monitoring:** OTMs closely observe and monitor the exercise as it unfolds, paying particular attention to the actions, decisions and performance of participants. This allows them to provide immediate feedback during the exercise, which can help participants adjust their approach to align with TOs.

³ Training objectives tend to be developed by participating nations and organisations and they vary in quality and format. In exercises with a large civilian dimension, challenges in finding common training objectives led on occasion to the absence of shared TOs.

2. **Data collection:** OTMs observe, analyse and report the TA performance based on the TO. They collect data and observations on various aspects of the exercise, such as communication, decision-making, leadership and overall execution. These data serve as a valuable source of information to assess whether the TA successfully performs established processes and procedures and delivers products according to SOPs.
3. **Feedback and debriefing:** OTMs engage in on-the-job mentoring with participants to facilitate the learning process and the fulfilment of TOs. These are followed by debriefing sessions designed to provide constructive feedback and insights, highlighting areas where TOs were met and improvements are needed. These take the form of a daily Hot Wash Ups and a final AAR.

OTMs compile observations daily and share their assessments with the Exercise director, the Exercise Control function and the EVAL function. They complement this work with debriefings such as Hot Wash Ups and AAR. During VK22, more than 110 OTMs participated in the exercise on the NATO mission side. This amounted to an average of 1 OTM for 6 trainees in the air component and mission Headquarters (HQ), 1 OTM for 9 trainees in the maritime component and 1 OTM for 13 trainees in a brigade. On top of this, a few embedded mentors were present in key positions to guide trainees.

Hot Wash Up sessions are informal and conducted on-site in the evening outside of the game.⁴ These sessions enable real-time feedback and quick discussions while the exercise is still fresh in the participants' minds (Hedlund and Österberg 2013). Hot Wash Ups are particularly valuable for addressing critical issues or immediate lessons learned during the exercise. They provide a platform for the rapid identification of deficiencies, the immediate correction of errors and the adjustment of tactics or procedures to ensure that TOs are met. These sessions are essential for enhancing the adaptability and real-time decision-making capabilities of military personnel, and thus play a significant role in the learning process and the fulfilment of TOs (Petranek et al. 1992).

4 Hot Wash/Hot Wash Ups. Conventional terms used to describe various ways in which Allied Command Operations (ACO) Commanders may conduct informal debriefings or follow-up discussions and evaluations of the performance of a HQ or multiple HQs during an exercise or major event or following its conclusion. The main purpose of a Hot Wash is to identify strengths and weaknesses recognised during the exercise/event, which may then lead to identify lessons in order to avoid repeated errors made in the past. A Hot Wash Up normally includes all the parties that participated in the exercise or event (NATO 2013: A-19).

AAR has been described as 'arguably one of the most successful organizational learning methods yet devised' (Senge 2001). Ellis and Davidi (2005, p. 857) define AAR as 'an organizational learning procedure that gives learners an opportunity to systematically analyze their behavior and to be able to evaluate the contributions of its various components to performance outcomes'. North Atlantic Treaty Organization (2013, p. A-7) defines AAR as a 'facilitated discussion that actively involves the TA. Through self-discovery, the TA will discuss the following three basic questions about performance in relation to the TO: What happened? Why did it happen? How can we do it better?'

The AAR is a structured and methodical session conducted immediately after the exercise (Mastaglio et al. 2011). It brings together all relevant participants, from military personnel to instructors and observers, to debrief and analyse the exercise performance. This review allows for the systematic examination of the execution of TOs. The AAR facilitates a discussion of what went well and what needs improvement. Participants are encouraged to provide feedback on various aspects, such as tactics, techniques, communication, leadership and the overall coordination of the exercise. The information gathered during AAR not only helps assess the fulfilment of TOs but also provides valuable insights for refining training methods and addressing any gaps. Highly structured forms of AAR have been found to be more effective than less structured forms in the military (Keiser and Arthur 2021).

Incorporating both AAR and Hot Wash Up sessions into the assessment process for TOs ensures that TA performance is evaluated comprehensively and with a focus on objectivity. These mechanisms are critical in providing feedback, identifying strengths, weaknesses and areas for improvement and ultimately contributing to the enhancement of training outcomes. Now that we have reviewed the evaluation framework used in CJSE and VK, we can turn to Kirkpatrick's Four-Level Model for training evaluation and explore how it can be applied to these staff exercises. This will enable us to pinpoint the gaps in the application of the framework which will open in turn the door for a more standardised process to assess the fulfilment of TOs.

4 Evaluation levels

Kirkpatrick's Four-Level Model is a widely used framework for evaluating training programmes' effectiveness and assessing the impact of learning interventions. Developed by Kirkpatrick (1976), this model consists of four

hierarchical levels, each focusing on a specific aspect of evaluation:

Level 1: Reaction measures participants' immediate reactions and responses to the training programme. It assesses their satisfaction, engagement and perceptions about the training content, materials, instructor and overall experience.

Level 2: Learning evaluates the extent to which participants have acquired new knowledge, skills or attitudes as a result of the training. It focuses on measuring the increase in knowledge, the development of new competencies and changes in behaviour due to the training.

Level 3: Behaviour assesses whether participants apply what they have learned in their workplace or actual job environments. It examines whether there is a noticeable transfer of skills, behaviours or changes in job performance resulting from the training.

Level 4: Results looks at the broader impact of the training on the organisation's goals and outcomes. It evaluates the overall impact of the training in terms of its contribution to organisational performance, productivity, cost-effectiveness, customer satisfaction and other key metrics.

Kirkpatrick's Four-Level Model provides a structured approach to evaluate training programmes comprehensively, starting from participants' immediate reactions, progressing through measuring learning outcomes and behavioural changes and culminating in assessing the training's impact on organisational results.

CJSE and VK evaluations cover the first three levels of the model. The fourth level falls outside the mandate, scope and timeframe of exercises and could be conducted by the Swedish Armed Forces headquarters.⁵ During staff exercises, the EVAL function and OTMs share the responsibility for evaluating various levels of Kirkpatrick's model. In practice, the division of labour is not very clear, and evaluations made by OTMs and EVAL members tend to overlap.

The evaluation of TA reaction (Level 1) has been informally done by OTMs during the first days of exercises to ensure that participants settle into their role, have access to important information and begin to produce according to the battle wheel and expectations. OTMs have done so to ensure that the grounds are laid properly before they can start to evaluate the fulfilment of TOs. Some elements from this informal information have been reported by OTMs, in their contribution to, for example, the Hot Wash

Ups and AAR. Formally, TA reaction has been evaluated by the EVAL function through the delivery of anonymous STARTEX and ENDEX surveys with a list of questions related to the exercise platform, IT system, scenario, Work Up Staff Training and professional development. Interviews with participants, especially commanders, as well as observations have been used by EVAL to collect data for their final report.

The evaluation of TA learning (level 2) has been formally done by the OTMs and to a lesser degree EVAL. Kirkpatrick and Kirkpatrick (2006, p. 22) define learning 'as the extent to which participants change attitudes, improve knowledge, and/or increase skill as a result of attending the [exercise]'. Learning is assumed to have taken place when attitudes, knowledge or skills have improved. TA learning has been at times evaluated by the EVAL function through the delivery of anonymous STARTEX and ENDEX surveys with a list of questions related to skills, knowledge and abilities. This enables the evaluation of progress by comparing the responses of the two surveys and identifying increases in scores. The questions are rather general in character and the answers are based on self-reporting, the limits of which have been identified in numerous studies (Paulhus and Vazire 2007). The bias could be increased by the fact that the 'peace operation' identity of exercise participants is tested first in the questionnaires (Koller et al. 2023).

Besides the surveys that provided a much-needed participant-focused perspective, learning was evaluated by the OTMs through daily evaluations and assessments as described previously. These are made using a grading scale based on Bloom's taxonomy and its three learning levels – remember/understand, apply and analyse/evaluate (Bloom et al. 1956; Krathwohl 2002). For example, during CJSE19, learning at level 1 according to Bloom's taxonomy expected participants to 'recognise basic knowledge of staff routines and processes/SOP in own cell/branch'. Learning at level 2 expected participants to 'apply intermediate skills in producing reports and orders'. Finally, learning at level 3 expected participants to 'assess advanced understanding of planning and conduct of a CRO in Joint Operations Area'.⁶ No grading scale was available to help this part of the evaluation process.

⁵ Since Swedish participants tend to participate in more than one such exercises, it is conceivable to conduct level 4 evaluation from one year to another for specific participants.

⁶ It should be noted that the description of the three levels of Bloom's taxonomy in the case of CJSE19 was very weak and inconsistent. For example, level 3 should concentrate on higher connective abilities such as analysis and evaluation and yet the first points under level 3 expected participants to 'prove advanced knowledge of staff routines and processes/SOP in own branch and knowledge of other branches within staff'. This is clearly an issue of understanding which naturally falls under level 1 rather than an issue of evaluation as expected at level 3.

The evaluation of TA behaviour (level 3) is also conducted within the frame of staff exercises. While Kirkpatrick planned such evaluation after trainees had left the classroom and returned to their job, staff exercises allow time for behavioural changes to take place and are thus relevant arenas to conduct level 3 evaluations (Kirkpatrick and Kirkpatrick 2006, p. 53). The grading scale used in CJSE and VK exercises thus evaluated changes in behaviour aiming to reach a state where '[t]he staff functions are fully trained, experienced and practiced...[and] require only basic level of supervision and very little mentoring... [and] display initiative, can act independently and together with other functions, in coordinated staff processes, in a dynamic fashion, under demanding conditions and under challenging time constraints'. Besides knowledge, OTMs will evaluate the capacity of participants to interact with their colleagues and contribute to their mission in accordance with the command structure, Battle Rhythm (BR) and applicable orders. During CJSE19 and VK22, a total of 1,130 and 1,432 observations were respectively recorded by OTMs to evaluate the progression made by participants towards the fulfilment of TOs. In addition, a total of 110 and 189 assessments were respectively written to describe in detail the progress made by specific staff.

5 Limits in the application of the evaluation framework

CJSE and VK exercises rely on set evaluation events, objectives, timing, criteria and methods (Simpson and Oser 2003, p. 27). The evaluation framework used finds support in much of the academic literature, and over the years it has been found to be useful. Small improvements and changes have been made from exercise to exercise to remedy shortcomings and fine-tune the process. In this article, we claim that to track general TA progress in a detailed and timely manner, a review of the evaluation framework is necessary to facilitate its full operationalisation through focused reporting and measurement.

An evaluation framework was in place, and yet a thorough analysis shows that its application was insufficient to track general TA progress in a detailed and timely manner.⁷ Indeed, this framework gave the impression that

learning was being evaluated but it was incomplete and not detailed enough. Evaluation criteria together with timing, benchmarks and standards were missing, leading to difficulties in the standardisation of measurements and their aggregation. More specifically, challenges associated with unclear evaluation objectives and criteria have emerged and will be discussed below. Let us look at each of these two elements of the evaluation framework in turn.

5.1 Objectives

CJSE and VK have EOs, the fulfilment of which is evaluated by the EVAL function. The focus is directly on the evaluation of the adequacy of the planning and structure of the exercise in enabling the TA to meet their TOs. The actual evaluation of TA fulfilment of the TOs is conducted by OTMs. However, the EOs outlined above are formulated in such a way that EVAL members were implicitly led to evaluate the fulfilment of TOs. For example, evaluators reported on the quality of commander's briefings or covered the fulfilment of TOs in their surveys. In practice, this meant that EVAL made observations that were outside of their mandate and should have been made by OTMs. This issue seemed to have affected disproportionately remote sites.

OTMs also took it upon themselves to assess the fulfilment of EOs even though this was not in their area of responsibility. OTMs systematically reported on the fulfilment of the EOs in their reports compiled in the VK22 FER. This was certainly facilitated by the fact that many TOs overlapped with EOs. This lack of clarity on the division of labour led to overlaps in the evaluation process. The issue has recently been addressed through the publication of the *Reference Guide on the EVAL Function in International Command-Post Exercise* (Gelot and Stewart 2023).

5.2 Criteria

Difficulties also emerged in the evaluation of TOs. For example, a TO that was regularly used for staff exercises read as follows: 'Conduct Mid-Term Planning including planning, execution and assessments coordinated with relevant actors in accordance with valid SOP, OPLAN/JCO/

⁷ The analysis developed in this article is based on exercise records (observations, assessments and reports) and does not cover all OTMs activities such as real-time mentoring and oral contributions to AAR and Hot Wash Ups. It is likely that these unrecorded contributions to the evaluation process addressed in parts the gaps identified below,

though not systematically. The main issue that we aim to address is the difficulty to measure TA progress in a standardised manner and the resulting difficulties to aggregate results. It is possible that individual OTM had a detailed understanding of specific units' progress towards the fulfilment of TOs. However, such assessments were not aggregated in a timely manner to provide an adequate overview.

FRAGO, relevant documents and CC'. This TO on mid-term planning contained a long list of tasks to be completed and that included:

- Coordinate and execute Midterm planning through Joint Coordination Board (JCB), Joint Coordination Board Working Group (JCBWG) and related working groups or supporting subprocesses.
- Coordinate and execute Joint Mid-Term planning.
- Produce and execute Air Operations Directive (AOD).
- Coordinate and synchronise midterm plans.
- Conduct supported/supporting interrelationships during planning for mid-term land operations.
- Develop plans for transition into execution.
- Develop plans that mitigate the effects of maritime operations.
- Coordinate maritime logistics with the Joint Logistic Support Group (JLSG) and maximise the use of allocated resources.
- Coordinate and execute midterm logistic operations.

While identifying key evaluation areas or products expected from trainees, this list of tasks did not cover the entirety of the mid-term planning process. Expectations are different for the different components, and certain units are not mentioned.

Moreover, the TO and associated tasks do not have SMART benchmarks or criteria to evaluate TA performance. There was insufficient clarity regarding the threshold expected for the fulfilment of TOs. For example, when should an OTM report that the midterm plans are sufficiently coordinated and synchronised? The lack of detailed, standardised and comprehensive TOs and tasks meant that OTMs had to rely on their judgement and criteria to operationalise and evaluate TO fulfilment. They repeatedly complained about this lack of guidance. As a result, observations submitted by OTMs were not sufficiently comprehensive, systematic and methodical. Instead, they were superficial, anecdotal and overly general. While this does not mean that the evaluation process, including mentoring and training, was insufficient, the lack of systematic reporting meant that it was difficult to aggregate observations to measure the TA fulfilment of TOs on a daily basis with any degree of precision.

The limits in the application of the current evaluation framework are amplified by additional issues. For example, CJSE and VK are distributed staff exercises, and communication can be challenging across sites. The issues identified above seemed to have been much more prominent in remote sites, especially those that do not have a long experience of evaluation in large staff exercises. The OTM organisation is temporary in nature and

thus suffers from a range of issues that affect new groups. Increased training prior to STARTEX, better communication and clearer instructions could have mitigated some of the issues identified.

The technical system used for the recording and aggregation of observations should also be mentioned. Its complexity and lack of user-friendliness meant that OTMs faced yet another obstacle to reporting observations. This may in part explain why only about one-third of OTMs at the fictitious NATO mission HQ and Land Component Command (LCC) HQ used the system actively. The system was also unable to aggregate observations to present TA progress on a daily basis, and no dashboard was available.

The last section of the article describes a more detailed approach to assess the fulfilment of TOs. It also describes concretely a TO evaluation process that facilitates reporting and measurement. Through the development of SMART TOs subdivided into STOs and MLOs, the section describes a standardised and aggregated evaluation framework suitable to support the timely decision-making process of the exercise director and exercise control centre.

6 Rethinking TO

Following NATO's lead on instructional analysis, we have rethought the focus and organisation of the TO evaluation process in a way that facilitates reporting and measurement (North Atlantic Treaty Organization 2015, pp. 42–47). The first step was to narrow-down the focus of TOs from general processes that were hard to evaluate in a SMART way to more specific processes aligned with SOPs. Joint functions were identified as key focus points for evaluation, including command and control, manoeuvre, intelligence, fires, sustainment, information, protection and CIMIC. The objective is to describe TOs in greater detail in line with the actual processes that are expected from the trainees. As such, TOs are not only described generally but also subdivided into STOs which are in turn subdivided into MLOs. The result is a list of aggregated TO, STO and MLO that are SMART.

If we take the example of intelligence as a joint function, and more specifically of generic Intelligence, Surveillance, Target Acquisition, and Reconnaissance (ISTAR) Task Force, a series of TOs can be outlined. For example, the first TO reads 'TO 1 – The commander and staff conduct the operations process for command and control (C2)'. The Mission Essential Task that is associated reads 'Employ the staff organization and procedures to optimize its ability to plan, conduct and support the operation

successfully according to a valid battle rhythm (BR) and SOPs'. The supporting and enabling tasks include:

1. Establish a Battle Rhythm for command and staff activities in accordance with higher HQ's decision cycle and battle rhythm.
2. Establish an information management plan (IMP) to support the process of collection, collation, storage, processing, dissemination and display of information to and from Brigade HQ as well as inside the HQ.
3. Review and update the SOP according to experiences from staff work and lessons identified.

Two STOs are listed under this TO, namely, STO 1.1 – COM ISTAR Task Force (TF) AND ISTAR HQ controls operations through the elements of control and STO 1.2 – COM ISTAR TF and ISTAR HQ applies the operational planning process. Finally, STOs are further specified into MLOs. For example, STO 1.1 is subdivided into four MLOs as follows:

- MLO 1.1.1 – ISTAR TF HQ provides direction to each subordinate/attached unit.
- Comment: MLO is achieved when every subordinate unit receives direction (either written or verbal) from ISTAR TF HQ when assigned a mission. MLO is achieved when COM ISTAR and ISTAR TF HQ consistently do so within a 72-h timeframe.
- MLO 1.1.2 – ISTAR TF HQ provides feedback to each subordinate/attached unit.
- Comment: MLO is achieved when each subordinate unit receives feedback (either written or verbal) from ISTAR TF HQ within 48 h after completing a mission. MLO is achieved when COM ISTAR and ISTAR TF HQ consistently provide feedback within a 48-h timeframe.
- MLO 1.1.3 – ISTAR TF HQ provides information to each subordinate/attached unit.
- Comment: MLO is achieved when each subordinate unit receives information (either written or verbal) from ISTAR TF HQ about the higher echelon's mission/intent, ISTAR TF's ongoing operations/mission/intent, adjacent units' missions/intent, supporting units' mission/intent and enemy forces within 48 h after completing a mission. MLO is achieved when COM ISTAR and ISTAR TF HQ consistently do so within a 48-h timeframe.
- MLO 1.1.4 – ISTAR TF HQ maintains communication with each subordinate/attached unit.
- Comment: MLO is achieved when ISTAR TF HQ establishes means and methods for maintaining communication (through technical or non-technical means) with each subordinate unit and attached forces and consistently does so over a 72-h timeframe.

MLOs follow a binary grading scale whether they are fulfilled or not; simplifying observation and monitoring, their fulfilment is recorded in a database. While a small margin of OTM judgement is required to assess if specific MLOs are met, this is reduced to a minimum due to the specification of timings and thresholds. When observations of MLOs are recorded in a simple database, they can be aggregated to assess the fulfilment of their parent STO and ultimately TO. For example, if MLO 1.1.1, MLO 1.1.2 and MLO 1.1.3 are assessed to be complete, STO 1.1 will be assessed to be 75% complete and TO 1 will be assessed to be 37.5% complete as long as the other MLOs falling under the TO are unfulfilled.⁸ When reporting on MLO fulfilment, OTMs can submit comments if necessary. They may also report observations on an MLO without having to mark it as complete. The weighting of each MLO and STO can be adjusted based on the desired impact; for instance, certain MLOs might need to be completed daily to achieve fulfilment.

This new framework for TOs fills many of the gaps identified above. The level of clarity is improved with detailed criteria such as timings, benchmarks, standards and thresholds for completion. The binary and aggregable nature of grading enables the standardisation of measurement and hence the live visualisation of the overall progress made by the TA towards the fulfilment of TOs. It goes to some length to develop performance metrics that are widespread in simpler exercises such as physical performance or shooting accuracy and that enable the quantification and comparison of the performance of units and individuals. Moreover, this added clarity is anticipated to decrease the risk of overlap with the evaluation conducted by the EVAL function.

The suggested refinement of the evaluation framework proves to be somewhat time-consuming during the preparatory stages of an exercise. Traditionally, exercises tend to proceed at a measured pace during the planning phase, accelerating into a faster tempo during execution. Despite the initial increase in workload associated with the introduction of this new framework, the anticipation is that it will streamline reporting processes, ultimately reducing the overall workload during the execution phase.

⁸ In this example, all MLOs are taken to be equally important and are given the same weight in the quantitative assessment of the performance of trainees. Numerical scores for the fulfilment of each MLO, STO and TO can be fine-tuned to reflect the importance of each MLO for the completion of a TO. In some cases, certain MLO may be critical to the fulfilment of the task while others may be less important. The numerical score can be adapted to reflect this situation.

Beyond providing a real-time overview of TO fulfilment, the framework is expected to alleviate the workload for OTMs, allowing for a more efficient allocation of time, particularly for mentoring activities. Additionally, there is the potential for a transformative impact on the pedagogical approach, with the prospect of a more streamlined OTM organisation and a greater integration of embedded mentors. Such a shift holds the promise of enhancing communication between evaluators, creating a more dynamic and collaborative evaluation environment.

Finally, it is worth mentioning that an in-house software-based solution has been developed to test and implement the revised evaluation framework. A Microsoft SharePoint environment has been tailored to enable the timely compilation and presentation of OTM observations during exercises. The digital platform enables OTMs to effortlessly input data through intuitive SharePoint forms. The automated grading system stands out as a key feature, as it calculates the fulfilment of TOs and STOs. This functionality allows for continuous monitoring of performance, providing graphic visuals to illustrate the overall progress made by the TA on a daily basis.

7 Conclusion

The collaboration between Sweden and Finland in training their armed forces through exercises like CJSE and VK has been pivotal in preparing officers for diverse operational environments. However, while these exercises have consistently demonstrated their utility in building staff competencies, the evaluation framework used to measure trainee performance was not sufficiently detailed.

This article has analysed the existing evaluation process, highlighting its strengths and pointing to limitations in its application. It has underscored the deficiencies in the evaluation objectives, criteria and communication among evaluators, which have hindered the generation of specific, timely and reliable observations of the TA's progress.

Moreover, this article has proposed a method to refine the evaluation framework. By introducing SMART and aggregable TOs subdivided into STOs and MLOs, it provided a structured and measurable assessment mechanism. This framework is designed to support the exercise director and exercise control centre in making informed and timely decisions by offering a more granular and aggregated evaluation process.

In essence, the suggested evaluation framework seeks to address the shortcomings of the existing system by providing a more standardised, comprehensive and precise

approach to measure trainee performance during CJSE and VK exercises. Implementing this approach holds the potential to enhance the overall effectiveness of these exercises, ensuring that officers are better equipped to navigate complex conflict scenarios, contribute meaningfully to crisis response operations and excel in UN peace-keeping missions.

References

- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (eds.). (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: Cognitive Domain*. New York, NY: David McKay.
- Depledge, D. (2020). Train where you expect to fight: Why military exercises have increased in the high north. *Scandinavian Journal of Military Studies*, 3(1), pp. 288-301. doi: 10.31374/sjms.64
- Doran, G. T. (1981). There's a S.M.A.R.T. way to write management's goals and objectives. *Management Review*, 70(11), pp. 35-36. Available at: <https://community.mis.temple.edu/mis0855002fall2015/files/2015/10/S.M.A.R.T-Way-Management-Review.pdf> [accessed 22 January 2025].
- Ellis, S., & Davidi, I. (2005). After-event reviews: Drawing lessons from successful and failed experience. *Journal of Applied Psychology*, 90(5), pp. 857-871. doi: 10.1037/0021-9010.90.5.857
- Enstad, K. (2022). Professional knowledge through wargames and exercises. *Scandinavian Journal of Military Studies*, 5(1), pp. 233-243. doi: 10.31374/sjms.130
- Gelot, L., & Stewart, H. (2023). Handbook on the EVAL Function in International Command-Post Exercises (CPX). Enköping: C5 Joint Warfare Centre.
- Gelot, L. (2019). Adapting a governmental training platform to simulate peace operations in the classroom. *Journal of Political Science Education*, 17(2), pp. 269-284. doi: 10.1080/15512169.2019.1621180
- Goldstein, I. L. (1986). *Training in organizations: Needs assessment, development, and evaluation*. Pacific Grove, CA: Brooks/Cole
- Hedlund, E., & Österberg, J. (2013). Team training, team learning, leadership and psychology safety: A study of team training and team learning behavior during a Swedish military staff exercise. *Sociology Mind*, 3(1), pp. 89-98. doi: 10.4236/sm.2013.31014
- Keiser, N. L., & Arthur, Jr., W. (2021). A meta-analysis of the effectiveness of the after-action review (or debrief) and factors that influence its effectiveness. *Journal of Applied Psychology*, 106(7), pp. 1007-1032. doi: 10.1037/apl0000821
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating Training Programs: The Four Levels*. San Francisco, CA: Berrett-Koehler Publishers.
- Kirkpatrick, D. L. (1976). Evaluation of training. In: Craig, R. L. (ed.), *Training and Development Handbook: A Guide to Human Resource Development*. New York, NY: McGraw-Hill, pp. 18-1-18-27.
- Koller, K., Pankowska, P. K., & Brick, C. (2023). Identifying bias in self-reported pro-environmental behavior. *Current Research in*

- Ecological and Social Psychology*, 4, p. 100087. doi: 10.1016/j.cresp.2022.100087
- Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, 78(2), pp. 311-328. doi: 10.1037/0021-9010.78.2.311
- Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory Into Practice*, 41(4), pp. 212-218. doi: 10.1207/s15430421tip4104_2
- Mastaglio, T., Wilkinson, J., Jones, P., Bliss, J., & Barnett, J. (2011). *Current Practice and Theoretical Foundations of the After Action Review. Technical Report 1290*. United States Army Research Institute for the Behavioural and Social Sciences.
- North Atlantic Treaty Organization. (2013). *Bi-Strategic Command (Bi-SC 075-003): Collective Training and Exercise Directive (CT&ED)*. Available at: https://www.coemed.org/files/Branches/DH/Files_01/bi-sc-75-3_final.pdf [accessed 13 February, 2024].
- North Atlantic Treaty Organization. (2015). *Bi-Strategic Command (Bi-SC 075-007): Education and Individual Training Directive (E&ITD)*. Available at: https://www.coemed.org/files/Branches/DH/Files_01/Bi-SC_75-7_NEW.pdf [accessed 13 February, 2024].
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. In: Robins, R. W., Fraley, R. C., & Krueger, R. F. (eds.), *Handbook of Research Methods in Personality Psychology*, Vol. 1. New York, NY: The Guilford Press, pp. 224-239.
- Petranek, F. C., Corey, S., & Black, R. (1992). Three levels of learning in simulations: Participating, debriefing, and journal writing. *Simulation & Gaming*, 23(2), pp. 174-185. doi: 10.1177/1046878192232005
- Roennfeldt, C. F. (2022). Foreword to the special issue on military exercises and wargaming in professional military education. *Scandinavian Journal of Military Studies*, 5(1), pp. 192-196. doi: 10.31374/sjms.168
- Senge, P. (2001). Introduction. In: Darling, M., & Perry, C. (eds.), *From Post-Mortem to Living Practice: An In-Depth Study of the Evolution of the After Action Review*. Boston, MA: Signet.
- Simpson, H., & Oser, R. L. (2003). Evaluating large-scale training simulations. *Military Psychology*, 15(1), pp. 25-40. doi: 10.1207/S15327876MP1501_03