

VALIDATION OF TRADITIONAL COMPETENCY-BASED ASSESSMENT AND EXPERIENCE-BASED AUTHORISATION FOR AIR TRAFFIC CONTROLLERS

Jana Meier[1], Marco-Michael Temme[1], Robert Hunger[1], Shruthi Shetty[1], Maria Hagl[1], Oliver Ohneiser[1] & Michael Finke[1]

[1]German Aerospace Center (DLR), Institute of Flight Guidance, Braunschweig, Germany

Abstract

Air traffic management (ATM) faces growing challenges in operational flexibility due to air traffic controller (ATCO) shortages and rigid endorsement systems. This paper presents the first human-in-the-loop (HITL) validation of an experience-based allocation concept developed within the SESAR3 IFAV3 project. The concept uses the Upper Airspace Skill-Level Index (UASLI) to predict transferable operational competencies from aggregated experience data, such as radar recordings, operational logs and training records. A six-day simulation campaign at the Maastricht Upper Area Control Centre (MUAC) was conducted to evaluate the concept under close-to-reality simulation conditions. Nineteen operational ATCOs managed both endorsed and non-endorsed sectors. Experience-based allocation outcomes were compared with assessments from MUAC's Competency Based Assessment System (CBAS), used to evaluate student ATCOs during on-the-job (OJT) training. Depending on whether CBAS self-assessments or assessor evaluations were used as the reference, and on the aggregation method applied, disagreement rates between CBAS-derived suitability assessments and allocation decisions ranged from 31.3% to 66.7%. Under the same configurations, potentially safety-relevant allocation decisions, defined as cases in which an ATCO was classified as suitable despite an insufficient CBAS outcome, ranged from 1.9% to 6.3%, assuming CBAS represents the ground truth of ATCO suitability. Analyses of objective operational data revealed no measurable negative environmental impact. Aircraft trajectories in unfamiliar sectors were even more efficient, showing improved straightness and fewer interrupted altitude changes. The findings indicate that CBAS alone may not be sufficient for evaluating transferable competencies and highlight the value of integrating objective performance metrics. Overall, the results support further refinement of the experience-based allocation concept, particularly concerning the calibration and further development of the UASLI.

Keywords: Air Traffic Control, Rostering, ATCO Flexibility, Competency Monitoring, Minimum Competency Level Prediction

1. Introduction

Air navigation service providers (ANSP) are increasingly confronted with the challenge of maintaining sufficient operational capacity whilst facing a persistent shortage of qualified air traffic controllers (ATCO) [1]. At the same time, the training and qualification of new ATCOs continue to be time-consuming and expensive. In addition, maintaining multiple unit endorsements requires regular training and operational exposure for ATCOs [2].

Current regulatory frameworks require ATCOs to operate only within sectors for which they hold a formal unit endorsement obtained through an on-the-job training (OJT) [3]. Although this system ensures a high level of operational safety, it also limits the flexibility of ATCO deployment within air traffic services units (ATSU). The maintenance of several endorsements further increases organisational complexity, as ATCOs must continuously accumulate sufficient operational hours and complete regular assessments for each unit endorsement. Consequently, the practical flexibility gains of multiple endorsements may diminish once endorsement maintenance itself consumes a significant

amount of operational working time.

To address these challenges, the SESAR research project “Increased flexibility of ATCO validations - V3” (IFAV3) has investigated concepts for more flexible and competency-based ATCO deployment [4]. Within these activities, an experience-based allocation concept was developed to estimate whether unit-endorsed ATCOs may safely operate sectors without a formal endorsement. In this context, allocation refers to the assignment of ATCOs to sectors. Previous publications focused primarily on the conceptual framework, mathematical modelling, operational assumptions and the development of the monitoring and prediction methodologies underlying the experience-based allocation concept [5][6].

Following the development of the concept, the next research step was to evaluate the operational feasibility of the experience-based allocation concept through validation under simulation conditions. For this purpose, a human-in-the-loop (HITL) validation campaign was conducted together with the Maastricht Upper Area Control Centre (MUAC) in late 2025. This paper presents a subset of the results from this validation activity, with a particular focus on comparing the predictions generated by the experience-based allocation concept with the assessments derived from MUAC’s Competency Based Assessment System (CBAS). The CBAS is currently used for evaluating ATCO trainees. The objective of the assessment is to investigate the extent to which the predicted allocation outcomes correspond to operational competency assessments. These assessments are conducted by unit-endorsed ATCOs in the roles of planning ATCOs and assessors, as well as by the participating executive ATCOs themselves.

In addition to the CBAS-based assessments, further objective operational data were recorded during the validation campaign. These include additional quantitative measures currently undergoing further analysis. First preliminary insights into these additional datasets are provided for potential environmental effects associated with the introduction of the concept.

This paper is structured as follows: Section 2 provides some insights into previous research on flexible ATCO deployment concepts and introduces the experience-based allocation concept, as well as the CBAS methodology used for validation. Section 3 describes the validation methodology, including participant characteristics and the experimental setup. Section 4 presents the validation results and analyses of the relationship between experience-based allocation outcomes and CBAS evaluations. It also provides first insights into the analysis of objective operational data recorded during the trials, including possible environmental effects associated with the introduction of the concept under investigation. Finally, Section 5 summarises the main findings and provides concluding remarks. The remaining sections contain the acknowledgement, contact information and copyright statement.

2. Background and Conceptual Framework

This section introduces the conceptual framework underlying the validation trials and subsequent analyses. It first reviews previous research on flexible ATCO deployment concepts and positions the experience-based allocation concept within this line of research. The experience-based allocation concept and its underlying calculation model are then presented. Finally, the CBAS used at MUAC is introduced, as it serves as the primary reference methodology for the validation.

2.1 Existing Work on Flexible ATCO Deployment Concepts

Transport domains are currently investigating concepts for increasing operational flexibility and automation. Examples include remote train operations (RTO) in the railway sector [7] and maritime traffic management concepts for coordinating vessel traffic in port approaches [8]. Within air traffic management (ATM), similar developments are being driven by increasing operational complexity and the continuing shortage of ATCOs across Europe [9]. More flexible deployment concepts for ATCOs are expected to improve resilience and cost efficiency in ATM service provision, while also supporting future operational concepts such as virtual centres [10] and dynamic sectorisation [11].

Consequently, the deployment of ATCOs across different operational environments has become an important research topic within the SESAR programme [12]. Research activities in this area have focused on identifying mechanisms through which operational competencies acquired in one environment can be transferred to other, operationally comparable contexts.

Early studies on IFAV examined the feasibility of extending ATCO deployment beyond traditionally endorsed sectors by identifying similarities between operational environments and analysing the corresponding training and competency requirements [13]. These investigations established the foundation for a more systematic assessment of transferable operational expertise.

Building on this work, subsequent research explored different strategies for flexible controller endorsement and deployment [14]. Particular attention was given to the identification of operational characteristics that influence transferability, including sector structures, traffic patterns, operational procedures, and task demands. The objective was to determine under which conditions existing competencies could be recognised in comparable operational contexts, thereby reducing the need for extensive additional training [15].

More recent research within the IFAV3 project expanded these approaches towards practical implementation and validation [16]. This included the investigation of methods to reduce OJT effort, support the maintenance of multiple endorsements and facilitate deployment across operationally comparable sectors through dedicated decision-support tools and adapted competency management approaches. These studies also highlighted the relevance of flexible ATCO deployment as an enabling element for future ATM concepts.

Building on this work, the following section introduces the experience-based allocation concept developed within the IFAV research activities. The concept aims to support a more flexible deployment of ATCOs to sectors by systematically assessing transferable operational competencies derived from previous experience data.

2.2 The Experience Based Allocation Concept

The proposed allocation process for assigning ATCOs to sectors is based on two complementary elements: Smart Controller Experience Monitoring and Minimum Required Competency Level Prediction. Its aim is to support flexible rostering of ATCOs within ATSUs. This means that already unit-endorsed ATCOs should be enabled to operate safely in sectors of another unit within the same ATSU, even without completing full OJT in the respective environment. This is achieved through the systematic assessment of transferable competencies that can be derived from previous operational duties, training records and simulation experience. Within this approach, experience is understood as a measurable dimension of competency and forms the basis of a data-driven allocation methodology.

The proposed monitoring element continuously assesses the competency profile of each ATCO on the basis of existing Air Traffic Management (ATM) data and ATCO work experience data, such as radar recordings, operational logs, simulator sessions and training records, to quantify the actual control experience an ATCO has recently acquired. The competencies considered are intentionally defined independently of any specific sector, making them transferable across different operational sectors. These competencies are translated into a corresponding set of indicators.

The second element predicts the minimum competency level required to operate a specific sector during a defined period of time. This prediction uses the same indicator framework as the monitoring process and evaluates expected traffic demand and sector complexity on the basis of inputs such as flight plans, weather forecasts, airspace use plans and sector characteristics. The use of a common indicator structure enables a direct comparison between the monitored competency level of an individual ATCO and the predicted competency requirement of a sector for a defined period of time.

The underlying mathematical model used to process experience data is the Upper Airspace Skill-Level Index (UASLI) [17]. The UASLI defines up to twenty-two skill parameters, each representing a measurable aspect of the operational situations experienced by an ATCO, and provides the connection between abstract competency and observable experience. The skill parameters are grouped into thematic categories, such as traffic-related aspects, sector-related influences or time-related factors. Each skill parameter is based on a range of input variables that are directly measurable, objective and derived from real or simulated operations.

Equivalent to the concept, the UASLI is also structured in two parts. First, it provides a method for estimating the expected capabilities of an ATCO in any given sector. This is achieved by evaluating the defined skill parameters (detailed in Section 2.2.1) and considering recent operational experience within a specified period of time. Second, the UASLI also defines a method for predicting the minimum competency level required for a given sector during a defined time period [18].

Each skill parameter results in a value between 0 and 1. For interpretation, in general a higher skill value indicates a higher level of experience. The strategy defines that an allocation is feasible if the available competency level of an ATCO meets or exceeds the predicted requirement. The resulting allocation is limited to the specific operational conditions and time period for which the competency requirements were predicted. If the available competency level of an ATCO falls below the predicted requirement, no allocation is possible. For this purpose, the UASLI provides two options of handling the skill parameters.

First, all skill parameters can be aggregated into a single overall index, resulting in an ATCO-specific competency index (ASI) and a sector-specific requirement index (SSI), which enables a rapid comparison for rostering purposes. Second, competency and requirements can be compared at the level of individual skill parameters, thereby increasing transparency and enabling a more detailed assessment of strengths and limitations. As the initial objective was to test the concept of experience-based ATCO-sector allocation, the second approach was used for validation purposes. The complete allocation process for rostering purposes is illustrated in Figure 1.

A key challenge of this process lies in the uncertainties associated with traffic forecasts. As actual traffic demand may differ from predicted values, the strict application of calculated competency thresholds could lead to overly conservative or unstable allocation decisions. To address this issue, the concept incorporates a Competency Requirement Flexibility (CRF) mechanism as a controlled buffer.

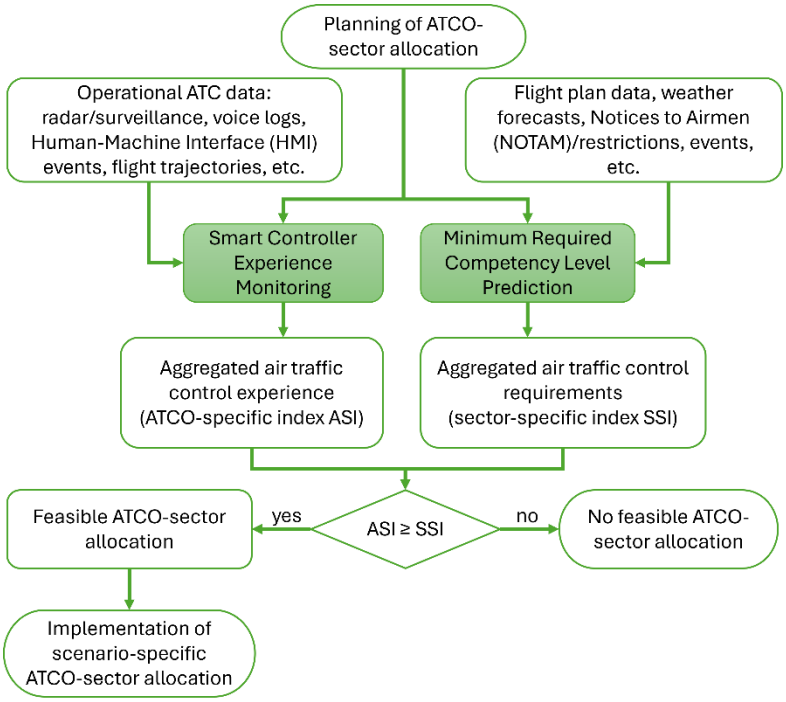


Figure 1 – Conceptual framework for experience-based ATCO-sector allocation.

The CRF introduces a positive adjustment margin based on selected parameters of the competency model. This margin can only increase the effective competency level and can never reduce it. In operational terms, this means that ATCOs may also be assigned to sectors where the predicted requirement slightly exceeds their nominal competency level. Based on expert discussions and ATCO workshops, the maximum flexibility margin was set at 10% of the competency value and may be applied individually to operational parameters.

2.2.1 Selection of UASLI Skill Parameters

Of the 22 skill parameters combined within the UASLI, 14 could be applied on the basis of the ATCO experience data available at MUAC. These parameters are listed in Table 1, together with a brief description of the expected impact on the skill level of an ATCO.

Table 1 – UASLI skill parameters applied in the trials and their expected impact on ATCO skill level.

Skill Parameter	Expected impact on ATCO skill level
Sector Occupation	A higher maximum load factor of a sector is assumed to increase operational workload and traffic complexity for ATCOs, therefore requiring a higher skill level.
Flight Level Diversity	The greater the flight level changes in the individual clearances, the greater the ability to cope in unfamiliar and more complex traffic situations.
Aircraft Type Diversity	As each aircraft type has different flight performance parameters, it can be assumed that a large number of different types also means a higher skill level in the coordination of heterogeneous air traffic.

Entry-Exit Altitude Differences	Due to the increased workload of synchronous horizontal and vertical movements of aircraft for ATCOs in sectors where this occurs frequently, a higher skill level can be assumed compared to ATCOs who only work in sectors where vertical movements occur only rarely.
Ratio of Vertical to Horizontal Movements	Vertical entries and exits make conflict detection and trajectory prediction significantly more complex, requiring ATCOs to maintain higher situational awareness and concentration levels than in predominantly horizontal traffic situations.
Intersection Complexity	A higher number of potential intersection points between aircraft trajectories is assumed to increase sector complexity and coordination workload. Consequently, ATCOs operating in environments with frequent potential conflicts are expected to require a higher skill level to ensure safe and efficient traffic separation.
Communication Events	Due to the increased workload associated with frequent radio and CPDLC communication events, a higher skill level can be assumed for ATCOs operating in communication-intensive traffic situations.
Radio Contact Times	Longer radio contact times are assumed to indicate a higher workload, as the ATCO is required to interact with a specific aircraft over an extended period of time.
Working Time	Working time refers to the actual time spent performing ATC duties. Longer operational working times are assumed to reflect a higher level of practical experience and routine in air traffic control tasks.
Military Airspace Restrictions	Frequent military airspace activities and temporary airspace restrictions are assumed to increase traffic complexity and coordination workload for ATCOs, resulting in higher operational demands and therefore a higher required skill level.
Weather Handling	Experience in severe weather situations is assumed to require a higher skill level, as adverse weather conditions increase traffic complexity and frequently necessitate tactical rerouting, conflict resolution and rapid coordination.
Sector Structure and Area	Sector structures with irregular shapes are assumed to increase operational complexity compared to sectors with a simple geometry. This may result in narrower or shorter traffic tracks, leading to a higher potential for conflicts.
Unit Endorsements	The more endorsements an ATCO already has for different sectors, the more experience she or he is considered to have with different traffic situations.
Professional Experience	Professional experience refers to the length of time an ATCO has held a valid unit endorsement and has been qualified to provide ATC services. Longer professional experience is assumed to reflect broader accumulated expertise and operational knowledge gained throughout an ATCO's career.

2.3 Competency Based Assessment System (CBAS)

The Competency Based Assessment System (CBAS) is used by the Maastricht Upper Area Control Centre (MUAC) to assess student ATCOs during OJT, with the aim to track progress towards obtaining a unit endorsement. Structurally, the system is comparable to other competency-based ATCO assessment frameworks described in the literature [19]. The application of the CBAS and the assessment of the ATCO are carried out by an instructor. The instructor is a unit endorsed ATCO with special training to allow supervision and training of students during OJT.

The CBAS is structured in a similar way to a questionnaire or checklist. In total, it comprises 41 questions in 11 competency categories. The categories group together questions relating, for example, to situational awareness, workload management or traffic management. Each question

offers the same response options: a) Significantly Below Standard, b) Slightly Below Standard, c) At Standard, d) Above Standard and e) Not Observed.

As part of the validation, the four assessment levels, a) to d), were converted into a numerical scale for the purpose of evaluation (Table 2). To remain consistent with the logic of the UASLI, where a higher value indicates a higher level of competency, the CBAS rating scale was likewise structured so that more positive evaluations corresponded to higher scores.

Only 28 of the 41 questions were considered for the validation, excluding questions related to the general qualification of becoming an ATCO. This approach was justified because all participants in the trials were fully qualified ATCOs holding at least one unit endorsement.

Table 2 – Conversion of descriptive CBAS rating levels into numerical equivalents.

Evaluation level	Numerical equivalent
Significantly Below Standard	0
Slightly Below Standard	1
At Standard	2
Above Standard	3
Not Observed	-

2.3.1 Aggregation Approaches for CBAS Ratings

In order to interpret CBAS outcomes appropriately, it is essential to understand how the overall CBAS ratings were derived. As CBAS was not originally designed to produce a single overall suitability rating, two aggregation methods were defined and applied for the purposes of this study: (1) the average value method and (2) the limit value method.

In the average value method, an overall CBAS rating is calculated as the mean of all numerical ratings assigned across the 28 assessment questions. This approach follows a compensatory logic, whereby lower ratings in individual categories can be offset by higher rating in others. Such an approach may be considered appropriate in the context of experienced ATCOs, where isolated weaknesses in less critical aspects should not necessarily prevent operational deployment.

In contrast, the limit value method does not rely on averaging but instead determines the overall CBAS rating based on the lowest rating observed among all assessment items. Under this approach, the overall CBAS rating is determined by the lowest rating assigned to any individual assessment question. Consequently, existing deficiencies cannot be compensated by higher ratings in other areas.

3. Validation Methodology

The validation was conducted in November 2025 in the MUAC training simulator environment with a HITL real-time simulation. The trials were carried out over a total period of six days.

The use of the MUAC simulation environment provided several advantages for the validation process. The controller working positions (CWP) used during the simulations were identical to those in the operational room, ensuring a highly realistic operational setting. As all participants were active MUAC ATCO staff members, no additional familiarisation with the simulation environment was required. Furthermore, the simulator environment enabled the development and adaptation of dedicated MUAC training scenarios specifically tailored to the requirements of the validation methodology.

3.1 Participant Characteristics

A total of 19 participants took part in the study. Among the participants, 16 were male and 3 were female. Participant age ranged from 24 to 55 years, with a mean age of 35.7 years and a median age of 38 years. All participants held two to three unit endorsements within all three sector groups: DECO, Brussels, and Hannover. Of these, 15 participants (78.9%) held two unit endorsements and 4 participants (21.1%) held three unit endorsements.

To ensure a balanced representation across the operational environment, an approximately even distribution of unit endorsements across the sector groups was targeted (Figure 2). The endorsements were distributed as follows: DECO East (n = 6), DECO West (n = 7), Brussels East (n = 8), Brussels West (n = 7), Hannover East (n = 7) and Hannover West (n = 7).

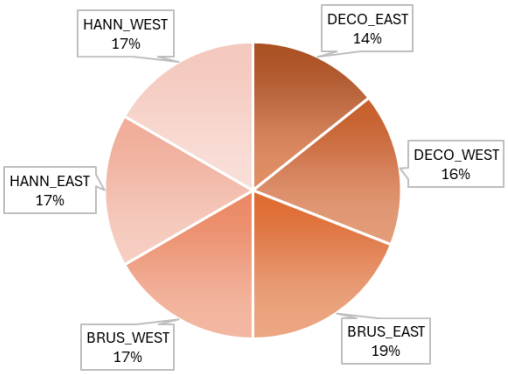


Figure 2 – Distribution of unit endorsements.

Operational experience of ATCOs involved ranged from 2 to 17 years. The mean level of experience was 8.95 years, with a standard deviation (SD) of 5.88 years. A clear gap in operational experience was observed between 7 and 11 years (Figure 3). The sample can therefore be divided into two subgroups: ATCOs with lower levels of experience and ATCOs with higher levels of experience. This distribution arose by chance, as participation was voluntary.

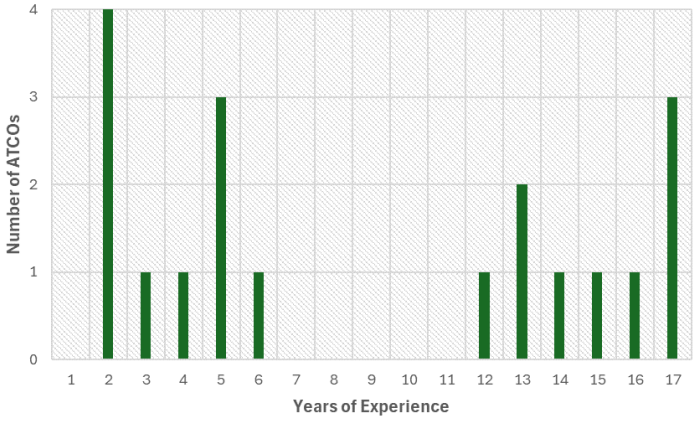


Figure 3 – Distribution of experience levels of participating ATCOs.

However, the relatively small sample size (n = 19) should be considered. Consequently, any subdivision into smaller groups should be interpreted with caution due to the limited number of participants in each subgroup. Whether and to which extent operational experience influenced the results is examined in Section 4.2.

3.2 Experimental Setup and Validation Procedure

The validation focused primarily on executive ATCOs responsible for the active management of traffic within the simulated sectors. Each executive ATCO was supported by a unit endorsed planner ATCO, who provided operational assistance whilst simultaneously monitoring and assessing the executive's performance throughout the scenario. Planner ATCOs were instructed to operate in the same manner as under live traffic conditions and to document interventions considered necessary due to potential concept-related deficiencies of the executive ATCOs. Depending on the complexity of the scenario and the surrounding traffic situation, adjacent sectors were additionally staffed by one or two operational ATCOs. To minimise potential assessment bias, planner ATCOs were assigned in a manner that avoided repeated assessment of the same executive ATCO across multiple scenarios wherever possible. In particular, no executive ATCO was assessed by the same planner ATCO in two directly successive scenarios.

Within the subsequent analysis, a distinction was made between self-assessments and assessor evaluations. The term self-assessment refers to the executive's own evaluation of their performance during a scenario. In contrast, assessors refer collectively to all ATCOs responsible for evaluating the executive ATCO. Both perspectives were analysed separately in order to identify potential differences between self-perception and external assessment.

For the trials, the sector selection was carried out by operational subject-matter experts from MUAC.

Three representative sector groups were chosen: Brussels, DECO and Hannover. Within each sector group, two operational sectors were selected. This resulted in a total of six sectors being included in the study: Luxembourg Total and West Total from the Brussels sector group, Delta Total and Holstein Total from the DECO sector group and Ruhr Total and East Total from the Hannover sector group (Figure 4). The selected sectors provided a representative sample of different traffic patterns and operational characteristics within the MUAC area of responsibility.

For each selected sector, two representative traffic scenarios were developed based on typical summer season traffic conditions. One scenario represented a comparatively moderate workload, whereas the second reflected a more demanding but still operationally realistic traffic situation. In addition, moderate-workload scenarios were modified to include turbulences and emergency-related events in order to evaluate ATCO performance under dynamically changing operational conditions.

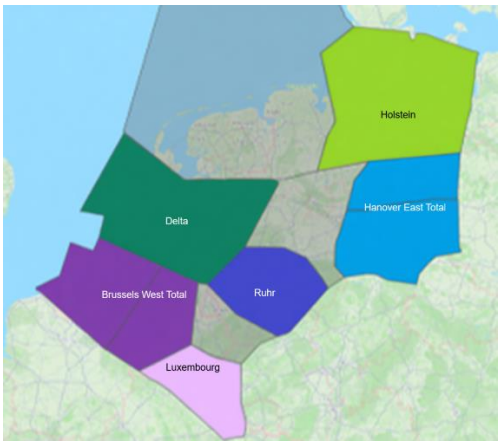


Figure 4 – MUAC sectors selected for the Human-in-the-Loop validation campaign. Two sectors were chosen from each of the three sector groups: Brussels, DECO and Hannover.

Each participating ATCO completed two reference runs in sectors for which a valid unit endorsement was held, followed by three solution runs in sectors where no endorsement existed. The reference runs were always conducted prior to the solution runs. To minimise potential learning and sequencing effects, the order of scenarios was varied according to operational constraints and support staff availability. Furthermore, different sectors were assigned across the solution runs wherever possible in order to reduce familiarity effects between scenarios.

For each participant, operational data recorded between August and October 2025 were compared with the predicted competency requirements of the respective simulation scenarios. Based on this comparison, scenarios representing both favourable and less favourable allocation outcomes were selected for validation purposes. If this selection allows, for each ATCO a moderate and a high traffic scenario was conducted. Furthermore, these two scenarios were selected to be conducted in the two different sectors of one sector group. The third non-nominal solution scenario as traffic wise identical with the moderate scenario were conducted in the other sector from the sector group the ATCO conducted their moderate scenario. The order between the three solution scenario types were shuffled individually per ATCO. Reference scenarios from individual ATCOs were subsequently compared with solution scenarios conducted by other participants in order to evaluate the consistency and validity of the experience-based allocation methodology.

3.3 Experience-Based Allocation Assessment

Prior to the trials, all solution scenarios were evaluated using the experience-based allocation methodology in order to determine whether the participating ATCOs were expected to be capable of operating the respective sectors or not. For this purpose, MUAC provided operational experience data for the 19 participating ATCOs covering the 90 days preceding the validation period. Based on these data, individual competency levels were calculated and subsequently compared with the predicted competency requirements of the solution scenarios.

Based on this comparison, the allocation outcomes were classified as either “able” or “not able”. Scenarios in which the monitored competency level met or exceeded the predicted sector requirement were categorised as “able”, whereas scenarios below the predicted requirement threshold were classified as “not able”. The allocation of ATCOs to scenarios was arranged in such a way that both positive and negative allocation outcomes were represented within the validation. As a result, 19

solution runs were classified as “able”, whereas 29 runs were assessed as “not able”, enabling the validation of the methodology in both directions. Across all participating executive ATCOs, a total of 48 valid solution runs were conducted.

4. Results

The following sections present the comparison between CBAS-based assessments and the experience-based allocation methodology, including the influence of operational experience, different aggregation approaches and the interpretation of observed discrepancies between both assessment concepts. In addition, selected analyses of objective operational data recorded during the trials are presented in order to investigate potential operational effects associated with the introduction of the experience-based allocation concept.

4.1 Statistical Analysis of CBAS Survey

To evaluate the prediction model, a shortened version of EUROCONTROL’s CBAS procedure was used, in which 28 of the total 41 questions were evaluated by participants in four different rules after each trial run. As part of the validation using reference and solution scenarios, a total of 7,303 questions regarding the 28 selected criteria were evaluated (“Sum evaluated” in Table 3). Of these, 2,169 received a rating of “Above Standard” (rounded to 29.7%), 4,708 were rated “At Standard” (64.5%), 379 were rated “Slightly Below Standard” (5.2%) and 47 were rated “Significantly Below Standard” (0.6%). For 89 (1.2%) responses, the evaluators were unable to observe a particular criterion and therefore felt unable to comment on it. The average rating on a scale from 0 (“Significantly Below Standard”) to 3 (“Above Standard”) was 2.23. As expected, the two reference scenarios – with scores of 2.52 and 2.51 – performed significantly better than the three solution scenarios, which scored 2.00, 2.05 and 2.07 on average.

Table 3 – Statistical summary of the ratings from the validation runs.

Scale	Value	Overall # complete	Executives # complete	Assessors # complete	Exec. & Assess. # complete
Above Standard	3	2169	526	1010	1536
At Standard	2	4708	1463	1291	2754
Slightly Below Standard	1	379	148	147	295
Significantly Below Standard	0	47	23	17	40
Not Observed	-	89	24	55	79
Sum evaluated (w/o Not Obs.)		7303	2160	2465	4625
Average		2.23	2.15	2.34	2.25

The subsequent analysis focuses exclusively on the assessors’ evaluations and the executives’ self-assessments. A total of 4,625 questions (2,160 plus 2,465) were analysed, with a score of 3 (rounded to 41.0%) given 1,536 times, a score of 2 (52.4%) given 2,754 times, 295 times a 1 (6.0%) and 40 times a 0 (0.7%). The average grade was accordingly 2.25, which is slightly above the average rating of all study participants.

The direct comparison of the assessors’ evaluations and the executives’ self-assessments is particularly interesting. The executives rated themselves much more critically with an average score of 2.15 than the assessors did in the same runs, with a score of 2.34 (Table 3). This is already evident in the executives’ reference scenarios, where they scored 2.46 compared to the assessors’ 2.74. However, the difference is particularly pronounced in the solution scenarios. The assessors rated all scenarios as “met on average” with scores of 2.02, 2.12 and 2.08, whereas the executives rated their own performance in the same trials as 1.90, 1.93 and 2.06 – with two scenarios smaller than 2.0 or evaluated as “not met.”

The distribution of assessor and executive ratings across all criteria is also interesting, as it reveals significantly different weightings or classifications of the demonstrated performance in specific situations and the CBAS criteria. In principle, it should be assumed that both a shortcoming and an outstanding achievement within a criterion are rated similarly in both self-assessment and external evaluation. However, the correlation diagram between assessor and executive ratings shows only an extremely weak correlation in the grading of individual performances (Figure 5).

Another striking finding is the significant concentration of “At Standard” (value of 2) ratings in the self-

assessment category, where the assessors rated the executives' performance in a much more nuanced (and, in most cases, higher) manner than was the case in the self-assessment. The evaluation discrepancy becomes particularly clear in the range above 2.6 on the y-axis, as in these scenarios the assessors rated the executives' performance as outstanding, whereas the executives themselves considered themselves only "average" or "as expected." The green line in Figure 5 illustrates this relationship once again: All points above this line indicate a higher performance rating by the assessors compared to the self-assessment, while all points below indicate a correspondingly more critical assessment of executives' performance. Assuming that the assessors' ratings are more impartial, only those ratings in which ATCOs rate themselves higher than the assessors did should be classified as critical. These four cases located in the lower-right quadrant of the diagram in Figure 5 are all close to the non-critical value of 2, and thus "At Standard."

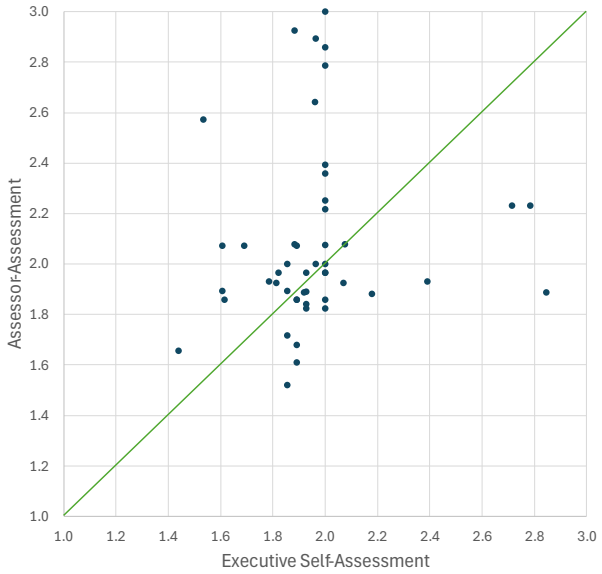


Figure 5 – Correlation between CBAS executive self-assessment and assessor assessment. Each data point represents an assessment of a traffic scenario. The green line indicates the ideal path if executives and assessors were to rate the trials identically.

However, CBAS was developed so that evaluators not involved in the simulation or traffic organisation could document and evaluate an objective assessment of an individual trainee performance. Its use as a self-assessment procedure thus appears to reach its limits in this specific validation and generate contradictions. This is particularly true for the IFAV3 trials, in which some of the younger assessors had never filled out a CBAS themselves before and may therefore have found it difficult to assign the performance they observed to a CBAS rating.

4.2 Influence of Operational Experience

The relationship between assessors and executives was examined with regard to their respective levels of operational experience. As illustrated in Figure 6, assessments were conducted across a wide range of experience combinations. These included cases in which assessors had greater experience than the executives, cases in which they had less experience and situations where both roles had comparable experience levels. Highly experienced ATCOs were evaluated by assessors with both high and lower levels of experience, likewise for less experienced ATCOs. No systematic pairing strategy based on experience alignment was identified. Instead, the distribution suggests a heterogeneous and balanced assignment approach, which reduces the potential for systematic bias related to authority gradients or shared experience levels.

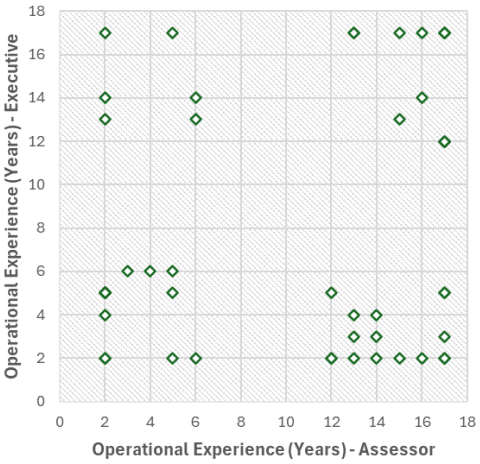


Figure 6 – Distribution of assessor-executive pairings by years of operational experience.

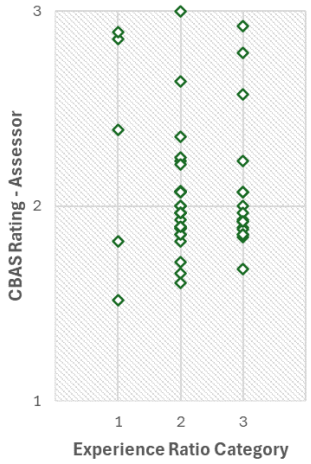


Figure 7 – Distribution of CBAS ratings across three assessor-executive experience ratio categories.

To further explore this aspect, a simplified categorisation was introduced to assess whether relative experience differences between assessor and executive influenced CBAS ratings. Only solution runs were considered for this analysis, and the CBAS ratings were derived using the average value aggregation method (introduced in Section 2.3.1). Three categories were defined: (1) assessor less experienced than the executive by more than 5 years, (2) assessor and executive with comparable experience (difference ≤ 5 years) and (3) assessor more experienced than the executive by more than 5 years. Across all categories, CBAS ratings largely overlap, with most scores clustering around the level “At Standard”, corresponding to a value of 2 (Figure 7). Both higher and lower ratings are observed in each category, and no consistent pattern emerges that would indicate systematically stricter or more tolerant assessments associated with any particular experience relationship. These findings suggest that assessor judgements are primarily driven by observed behaviour within the scenario rather than by relative experience differences.

4.3 Comparative Analysis of CBAS Results across Aggregation Methods

The comparative analysis was conducted to examine the extent to which CBAS-based results align with the “able” and “not able” classifications derived from the newly introduced experience-based allocation approach. The analysis was performed across all solution runs and included both CBAS self-assessments and assessor-based ratings. For each dataset, results were analysed separately using the average value method and the limit value method in order to assess the influence of the respective aggregation logic on the level of agreement.

To ensure comparability with the experience-based assessment approach, method-specific thresholds were defined for the CBAS ratings. For the average value method, CBAS scores above 1.9 were classified as acceptable, reflecting the smoothing effect of averaging and allowing values close to the “At Standard” level to be considered compliant. In contrast, the limit value method applies a stricter criterion, whereby a threshold of exactly 2 is required for a case to be classified as acceptable.

Furthermore, two types of false rates were considered. First, the false rate captures all cases in which the experience-based assessment does not correspond to the CBAS result. This includes instances where CBAS values fall below the defined threshold (i.e. <1.9 or <2, depending on the method) while the case is classified as “able”, as well as cases where CBAS values meet or exceed the threshold (≥1.9 or ≥2) but ATCOs are classified as “not able”. In addition, a critical false rate was defined, focusing exclusively on safety-relevant mismatches. This category includes only those cases in which CBAS values fall below the threshold while the experience-based approach classifies the outcome as “able”. This distinction is helpful, as not all mismatches carry the same operational relevance. While the overall false rate reflects general disagreement between the two approaches, the critical false rate specifically captures cases with potential safety implications, assuming ground truth captured by CBAS ratings.

4.3.1 Analysis of CBAS Self-Assessment Results

The analysis of CBAS self-assessments reveals a clear influence of the applied aggregation method on the level of agreement with the experience-based allocation outcomes. When using the average value method, 19 cases show agreement between CBAS-based classification and the corresponding “able” or “not able” decision, while 29 cases are classified as mismatches. This results in a comparatively high false rate of 60.42%, indicating a substantial level of disagreement between the two approaches under compensatory aggregation conditions (Figure 8).

In contrast, the application of the limit value method leads to a marked improvement in overall agreement. The number of matching cases increases to 33, while mismatches decrease to 15, resulting in a significantly lower false rate of 31.25%. This reduction suggests that the stricter, non-compensatory aggregation logic produces CBAS outcomes that are more consistent with the experience-based allocation decisions.

A different pattern emerges when considering safety-relevant discrepancies. The critical false rate increases from 4.2% under the average value method to 6.3% under the limit value method. Although the absolute number of such cases remains low, the stricter aggregation approach slightly increases the number of ATCOs classified as “able” despite an insufficient CBAS outcome. At the same time, the overall alignment between both assessment methods improves.

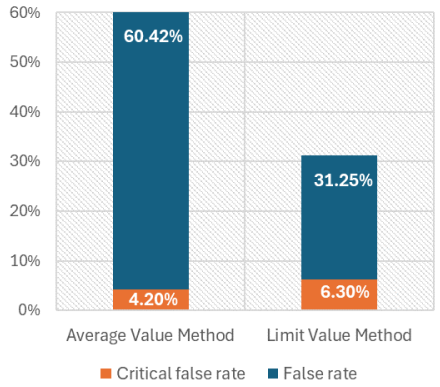


Figure 8 – Comparison of CBAS self-assessment results and experience-based allocation outcomes using the average value and limit value aggregation methods.

4.3.2 Analysis of CBAS Assessor-Based Results

The assessor-based evaluation shows a similar, but more pronounced, dependency on the aggregation method. Using the average value method, only 18 cases are classified consistently with the experience-based allocation, while 36 cases represent mismatches. This corresponds to a false rate of 66.67%, indicating an even lower level of agreement compared to self-assessments under the same method (Figure 9).

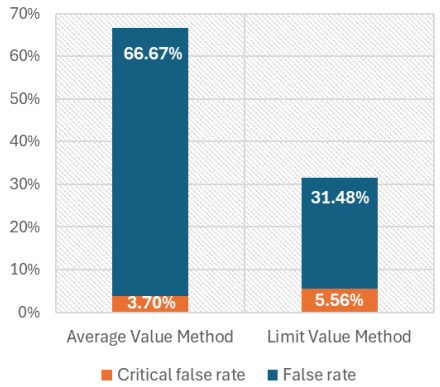


Figure 9 – Comparison of assessor-based CBAS results and experience-based allocation outcomes using the Average Value and Limit Value aggregation methods.

When applying the limit value method, the number of consistent classifications increases substantially to 37, while mismatches are reduced to 17. The resulting false rate of 31.48% is nearly identical to that observed for self-assessments, suggesting that the limit value method leads to a convergence of results across different rating sources.

With respect to safety-critical discrepancies, the critical false rate rises from 3.7% under the average value method to 5.56% under the limit value method. As observed for self-assessments, this reflects a slight increase in safety-relevant mismatches when applying the stricter aggregation rule. However, the overall level remains low, suggesting that such cases are relatively rare within the dataset.

In summary, the assessor-based results confirm the general pattern identified in the self-assessment analysis. The limit value method substantially improves overall agreement with the experience-based allocation approach, while simultaneously introducing a modest increase in critical false classifications.

4.4 Question-Level Analysis of CBAS Results

To complement the run-based analyses presented in Section 4.1and 4.3, the CBAS results were additionally examined at question level. Whereas the previous analyses aggregated the ratings of all 28 evaluated questions into a single CBAS outcome for each run, the present analysis investigates how individual questions were rated across all self-assessments and assessor-based evaluations.

For each of the 28 questions, mean ratings were calculated separately for executive ATCO self-assessments and assessor evaluations. Figure 10 presents the resulting average ratings. Overall, a strong correspondence can be observed between both assessment perspectives. Questions that received comparatively high ratings in the self-assessments were generally also rated highly by the assessors, whilst questions receiving lower ratings showed a similar pattern in both groups. This observation contrasts with the weak agreement observed in the run-based analyses (Figure 5). Whilst assessors and executive ATCOs often differed in their evaluation of individual simulation runs, both groups showed a remarkably similar pattern when the ratings were aggregated at question level. The results therefore suggest that assessors and executive ATCOs broadly agree on which aspects of performance are comparatively strong or weak, even if they differ in their assessment of individual operational situations.

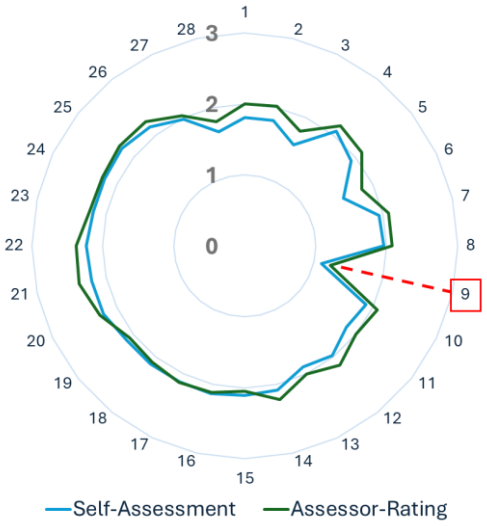


Figure 10 – Mean CBAS ratings (0 to 3 scale) by assessment criterion (1 to 28) for self-assessment and assessor evaluation. Lower ratings for question 9 indicate reduced procedural knowledge assessments.

The majority of the CBAS questions were rated around or above the “At Standard” level, corresponding to a value of 2. At the same time, the analysis reveals that several questions exhibit average ratings below that level. The most pronounced deviation can be observed for question 9, which received mean ratings of approximately 1.1 in the self-assessments and 1.2 in the assessor evaluations. Among all evaluated questions, this question therefore represents the clearest and most consistent deviation from the expected performance level.

The criterion represented by this question assesses the procedural knowledge of ATCOs within a respective sector or scenario. The strong alignment between self and assessor ratings suggests that the observed deviation is unlikely to result from subjective judgement. A plausible explanation for this finding can be found in the conditions under which the validation was conducted. The underlying concept assumes that ATCOs receive a short sector-specific briefing or familiarisation prior to operating within a given environment. Owing to time limitations during the validation phase, this preparatory step could not be carried out. Consequently, sector-specific procedural knowledge could

not be expected to reach the same level as the remaining competency dimensions, which is consistently reflected in the corresponding CBAS ratings.

Given this background, an additional run-based analysis was conducted in which the procedural knowledge criterion was excluded from the evaluation. The aim was not to obscure weaknesses, but to assess the extent to which a clearly context-driven outlier influences the overall alignment between CBAS and experience-based allocation outcomes.

For the self-assessment evaluations, the exclusion of the procedural knowledge criterion led to only minor changes when applying the average value method. The number of matching cases decreased slightly from 19 to 18, whilst the false rate increased from 60.4% to 62.5%. At the same time, the critical false rate decreased from 4.2% to 2.1%. In contrast, the limit value method showed a reduction in matching cases from 33 to 30 and an increase in the false rate from 31.3% to 37.5%, whilst the critical false rate remained unchanged at 6.3%.

For the assessor-based evaluations, the exclusion of procedural knowledge had no effect on the average value method in terms of matching cases and overall false rate. The number of agreements remained at 18 and the false rate remained unchanged at 66.7%. However, the critical false rate decreased from 3.7% to 1.9%. Within the limit value method, the number of agreements decreased from 37 to 29, whilst the false rate increased from 31.5% to 46.3%. The critical false rate remained unchanged at 5.6%.

Overall, the results indicate that excluding the procedural knowledge criterion did not improve the overall agreement between CBAS assessments and the experience-based allocation model. However, the exclusion resulted in a substantial reduction in critical false rates. Table 4 summarises the results of all evaluated configurations, including self-assessment and assessor-based evaluations, applying the average and limit value method, both with and without the procedural knowledge criterion.

Table 4 – Summary of results for the comparison between CBAS assessments and experience-based allocation outcomes across all evaluated configurations.

Evaluation Type	Configuration	False rate [%]	Critical false rate [%]
Self-Assessment	Average value	60.4	4.2
	Limit value	31.3	6.3
	Average value (without procedural knowledge)	62.5	2.1
	Limit value (without procedural knowledge)	37.5	6.3
Assessor-Based	Average value	66.7	3.7
	Limit value	31.5	5.6
	Average value (without procedural knowledge)	66.7	1.9
	Limit value (without procedural knowledge)	46.3	5.6

Across all evaluated configurations, a substantial discrepancy can be observed between the overall false rates and the critical false rates. Whilst the total number of mismatches is comparatively high, the proportion of safety-critical mismatches remains consistently low. This indicates that the experience-based allocation process, and in particular the underlying UASLI calculation, tends to behave conservatively. In practical terms, the model appears more likely to reject potentially feasible allocations than to permit allocations that could result in safety-critical situations.

4.5 Discussion of CBAS and Experience-Based Allocation Discrepancies

The presented results revealed a substantial discrepancy between CBAS results and the assignments generated by the experience-based assessment. Several factors may have contributed to these deviations and may have limited the direct comparability between both approaches.

First, the CBAS was not applied under the conditions for which it was originally designed. The system was developed to support the assessment of student ATCOs throughout the OJT process, where repeated observations over weeks or months contribute to the overall evaluation and progression towards a unit endorsement. In the trials, however, each executive ATCO was typically assessed in only three solution runs. Furthermore, the solution runs intentionally involved sectors for which the participating ATCOs did not hold a unit endorsement. As a result, the CBAS outcomes represent isolated observations in an unfamiliar operational context. This difference in both temporal scope and operational context may have limited the direct comparability between the two approaches.

A related observation can be derived from the comparison between assessor ratings and executive self-assessments presented in Section 4.1. The correlation between both assessment perspectives was found to be extremely weak, whereas executives generally rated their own performance lower than the assessors. One possible explanation is that executives did not evaluate their performance solely against the requirements of the specific traffic situation observed during the simulation. Instead, they may have compared their performance against their own professional expectations and routine operational standards developed within sectors for which they hold a unit endorsement.

A further distinction can be observed in the scaling structure of both methodologies. CBAS is based on a qualitative ordinal rating scale using four behavioural performance categories. The UASLI model, however, evaluates up to twenty-two individual skill parameters, each of which can assume continuous values between 0 and 1. Consequently, the UASLI does not rely on discrete assessment levels, but instead uses a continuous scale with gradual transitions between lower and higher competency levels. As a result, both approaches provide different forms of information and may react differently to variations in operational performance. Certain operational effects addressed by the UASLI are not necessarily observable or meaningfully assessable within a single real-time behavioural assessment situation. Conversely, CBAS is specifically designed to capture immediate behavioural and operational performance.

An important observation can be derived from the reference runs. Even in scenarios where ATCOs operated within sectors for which they held a valid unit endorsement, individual CBAS questions occasionally received ratings below the “At Standard” level (Section 4.1). This observation indicates that lower ratings in individual questions do not necessarily imply insufficient operational performance or an inability to safely operate the sector. Rather, deviations from the expected rating level may occur even in familiar operational environments. One possible explanation is that individual CBAS questions can be influenced by isolated observations occurring during a simulation run. For example, a temporary uncertainty regarding a particular traffic situation may result in a lower rating for a specific question, even though the overall traffic situation remains safely and efficiently managed. Such observations are highly relevant from a training perspective and therefore fully justified within the CBAS framework. However, they do not necessarily indicate a reduction in overall operational capability. Consequently, the occurrence of false and critical false cases should not be interpreted under the assumption that CBAS represents an error-free reference. Instead, the results raise the question of what level of disagreement between both approaches can realistically be expected.

The analysis presented in Section 4.4 further suggests that not all CBAS questions necessarily contribute equally to the operational objective addressed by the experience-based allocation concept. The procedural knowledge question provides a particularly illustrative example. The experience-based allocation concept was not designed to reproduce all dimensions captured by the CBAS. Rather, the underlying UASLI model focuses on predicting whether an ATCO can safely and efficiently operate a sector under specific operational conditions.

Finally, several methodological limitations of the validation setup may have influenced the comparability between both approaches. These include the relatively small sample size and the absence of comprehensive sector familiarisation prior to the validation runs. In operational practice, ATCOs would normally receive a structured briefing or familiarisation period before operating within an unfamiliar sector environment.

In addition to the CBAS-based assessment, further possibilities exist to evaluate the experience-based allocation methodology and, in particular, the functionality of the UASLI model. During the trials, a broad range of operational data was recorded within the MUAC simulator and made available for subsequent analysis. These data include, for example, flight path data, communication-related data and additional operational system parameters. Such data sources enable the analysis of operational effects that may not be fully reflected within behavioural assessment methodologies alone. Section 4.6 therefore presents initial analyses of these objective simulation data with regard to possible environmental influences associated with the introduction of the new allocation concept.

4.6 Assessment of Environmental Impact

As part of the validation activities, several aspects associated with the introduction of the experience-based allocation concept were investigated, including safety, human performance, workload and potential environmental effects. To support these analyses, a broad range of operational data was collected during the simulation trials. The environmental assessment presented in this section represents one exemplary subset of the overall validation data analysis. For this purpose, several

operational parameters derived from flight paths and altitude management were analysed and compared between reference and solution scenarios.

4.6.1 Assessment Parameters

Within the scope of the validation exercises, no significant impact on fuel efficiency or emissions was expected, as the experience-based allocation concept does not introduce changes to aircraft trajectories, flight profiles or airspace structures. The purpose of the assessment was therefore to verify whether the introduction of the solution could lead to any unintended environmental effects.

Due to the test design and the data availability from the validation runs, it was not possible to calculate or estimate the direct fuel consumption and thus also the CO_2 emissions of the aircraft. Aircraft in upper airspace typically have relatively constant fuel consumption, which depends on the aircraft type, current weight, flight level, air temperature and several other parameters. Fuel consumption is greatly influenced by the length of the route and by interrupted climb and descent profiles requiring repeated engine throttling and thus increased fuel consumption. This means that the usable parameters are straightness, which describes the ratio between the distance flown and the direct distance between a sector's entry and exit points, and the number of level-offs, which indicate intermediate flight levels assigned by ATCOs before another level change.

The *Straightness Ratio* indicates how straight the flight path was in a given airspace, taking into account the entry and exit points of the operational sector in relation to the route planned in the flight plan. A straightness of 1.0 describes a path directly along the global great circle and thus represents the shortest connection between two points on a sphere (1.0 thus describes the ideal connection). A straightness of 1.1 therefore corresponds to a route that is 10% longer than the theoretically shortest connection between origin and destination.

The parameters available from the simulation systems allow MUAC the calculation of the great circle connections between planned entry and exit points in sectors, the calculation of the length of the route based on the flight plans, and the length of the actual distances traveled by aircraft based on the guidance of the ATCOs. If the straightness of the reference scenarios is the same or at least similar to that of the solution scenarios, this can be considered a strong indicator of low additional environmental impact from the exercise of unfamiliar sectors. If ATCOs have difficulty guiding traffic safely and efficiently in an airspace, this is likely to result in additional turns due to evasive maneuvers, which further distances the straightness in the solution scenarios from the ideal route.

The second parameter *Intermediate Flight Level* (IFL) can be used as a measure of the foresight and workload of an ATCO in a sector. During a flight, aircraft must regularly change flight level. There are many reasons for this, but most often it is to climb to cruising level, descend when approaching the destination airport or change level to optimize fuel flow or avoid potential conflict situations. For safety reasons especially in the vicinity of airports, these level clearances are often given in several small steps. If there are no significant traffic-related reasons, the next flight levels are cleared before the last cleared level is reached. By avoiding horizontal intermediate flight segments (level flights), fuel consumption can be reduced during the entire climb or descent. If an ATCO misses the timely clearance of the next flight level, this indicates inattention due to overload or a lack of familiarity with local procedures, thereby increasing the environmental impact of a flight.

4.6.2 Environmental Results

A total of 83 scenarios involving 4,394 aircraft were recorded, for each of which a complete data set is available. Of these, 1,907 aircraft were included in the reference scenarios and 2,487 aircraft were part of the solution scenarios. The solution scenarios differed in terms of the participants who were predicted to be able to safely manage the traffic ("able", 913 aircraft) and those for whom the traffic situation was predicted to pose a particular challenge ("not able", 1,574 aircraft). The environmental assessment now refers exclusively to the reference scenarios and the solution scenarios of the "able" ATCOs, with a total of 2,820 flight movements evaluated.

The Parameter *Straightness*

The average straightness (unitless) based on pre-simulation flight plans was 1.011 in reference scenarios, which was only a little over one percent above the ideal routes along the great circle (Figure 11). In the corresponding refence simulations, the pilots actually flew their aircraft even more directly, achieving a straightness of 1.009 and deviating a little less than one percent from the ideal route. In the solution scenarios, the straightness based on the flight plans was also 1.011. The straightness

achieved by the ATCOs was subsequently 1.010, which was on average 1.33% better than the flight plan routes. This means that, with an average flight distance through a sector of 150 km, there was an average route shortening of around 2,000 meters compared to the flight plans. Broken down, 712 aircraft were guided on a shorter route and only 201 on a longer route than specified in the flight plans.

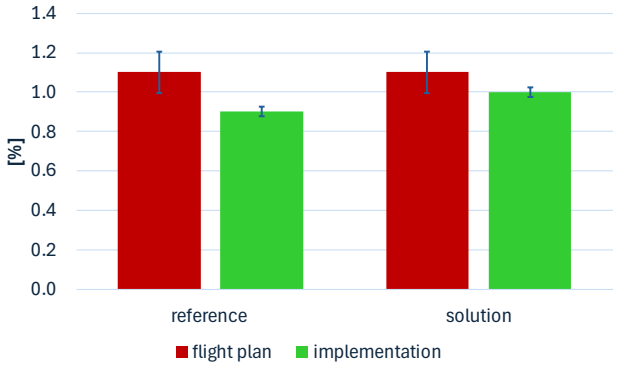


Figure 11 – Comparison of the average straightness improvement in the reference and the solution scenarios. The y-axis indicates the percentage by which the flight routes were extended compared to the great circle route. In both scenarios, the ATCOs managed to guide the aircraft along shorter routes than those specified in the flight plans (lower is better).

The standard deviation of the percentage route differences is interesting in this context. This was 10.72% in the references, but fell to surprising 5.28% in the solutions. On the one hand, this has a direct impact on the environment, as it shows that there were fewer "deviations" from the planned route than in the reference scenarios. On the other hand, it indicates that ATCOs in a sector previously unknown to them were very focused on always guiding traffic along the ideal line between the sector entry and exit points. This meant that fewer additional maneuvers in flight were required to manage traffic safely and efficiently in the unknown sectors.

The Parameter *Interrupted Altitude Changes*

As part of the traffic simulations, data was also collected on whether an aircraft had already been cleared for further level changes before reaching its cleared level. This ideally prevents aircraft from having to transition to level flight in the meantime. Most flights did not require intermediate altitudes to be flown, but there were slight differences between the reference scenarios and the solutions. In the reference scenarios, pilots were able to dispense with level off in 88.6% of cases (1,690 flights) and in the solutions this figure was as high as 91.6% (836 flights) (Figure 12).

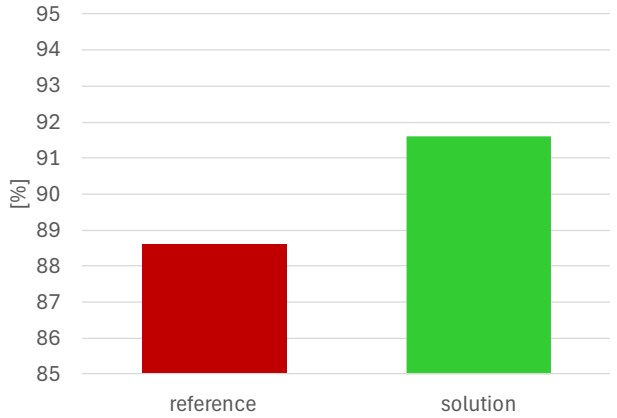


Figure 12 – Comparison of the sum of uninterrupted flight altitude changes in relation to all aircraft in percent. It is clear that ATCOs were able to guide aircraft to their target altitude without additional level segments more frequently in the solutions than in the reference scenarios (higher is better).

Accordingly, single level offs in the solutions only occurred in 7.8% of cases, compared to 10.5% in the references. A double level off occurred in 0.55% (5 aircraft) of the solutions, compared to 0.79% (15 aircraft) in the references. A triple interruption of a climb or descent occurred once in both the references and the solutions. It is therefore not possible to make a statistical statement about triple interruptions. On average, ATCOs guided aircraft slightly better and thus more environmentally

friendly during altitude changes in the solution scenarios than in the reference scenarios.

Overall, the results from the HITL simulations confirm that experience-based ATCO-sector allocations do not demonstrate a measurable or systematic impact on fuel efficiency or emissions, supporting the initial assumption of environmental neutrality for the concept.

5. Conclusion

Existing ATCO endorsement schemes provide a robust framework for ensuring safe ATC operations but inherently restrict the flexible deployment of ATCOs across sectors. The experience-based allocation concept investigated in this study addresses this limitation by introducing a data-driven approach for estimating transferable operational competencies and supporting ATCO-sector allocation decisions beyond formal endorsement boundaries.

The primary objective of this study was to investigate whether such an experience-based allocation concept can provide meaningful allocation recommendations for ATCOs operating in sectors for which they do not hold a formal unit endorsement. For this purpose, a HITL simulation campaign was conducted at MUAC in the Netherlands. During the simulations, 19 volunteering ATCOs operated both sectors for which they already held a unit endorsement and sectors unfamiliar to them. Prior to the trials, the experience-based allocation concept classified these unfamiliar-sector allocations as either feasible (“able”) or not feasible (“not able”).

The initial analyses focused on the potential influence of operational experience on the behavioural assessments conducted during the trials. No systematic influence of relative operational experience between assessors and executives could be identified. The results suggest that assessor evaluations were primarily driven by the observed operational behaviour within the scenarios rather than by experience-related bias.

Afterward, the experience-based allocation outcomes were analysed. The validation combined the experience-based allocation results with behavioural performance assessments using MUAC’s CBAS, which is applied during OJT training to evaluate student ATCO performance. The comparison between both revealed substantial differences depending on the applied evaluation approach. Two aggregation methods were investigated: an average value method, allowing strengths in some assessment areas to compensate for weaknesses in others, and a limit value method, where the lowest-rated question determined the overall CBAS outcome. In addition, both executive self-assessments and assessor-based evaluations were considered.

Depending on the applied evaluation approach, the comparison between CBAS assessments and the allocation outcomes resulted in false rates between 31.3% and 66.7%. At the same time, the corresponding critical false rates, representing cases in which an ATCO was classified as suitable by the experience-based allocation concept despite an insufficient CBAS outcome, remained consistently low, ranging only between 1.9% and 6.3%. This substantial difference between overall mismatches and safety-critical mismatches suggests that the experience-based allocation concept, and in particular the underlying UASLI model, currently follows a comparatively conservative allocation logic. In practical terms, the concept appears more likely to reject potentially feasible allocations than to permit allocations that could result in operationally critical situations.

The analysis further highlighted notable differences between assessor-based CBAS evaluations and executive self-assessments. Whilst both assessment groups rated the overall operational performance above the “At Standard” (2) level on average, executives generally assessed their own performance more critically than the external assessors. Across all analysed scenarios, the average self-assessment score was 2.2, compared to 2.3 for the assessor evaluations. In addition, the correlation between self-assessments and assessor evaluations was found to be weak. Self-assessment ratings showed a strong concentration around the “At Standard” category, whereas assessor evaluations differentiated operational performance more clearly across the full rating scale.

A possible explanation discussed in this study is that executives evaluated their performance against internal professional benchmarks and routine operational standards derived from sectors in which they normally operate, rather than solely against the requirements of the individual simulation scenario. Factors such as uncertainty, increased mental workload or reduced confidence when operating in unfamiliar sectors may therefore have influenced the self-assessments. However, a generally more self-critical assessment tendency was also observed within the reference scenarios. Consequently, the observed differences between self-assessment and external assessment cannot be explained exclusively by the unfamiliar-sector conditions.

In addition to the behavioural assessments, the validation also enabled first analyses of objective

operational data recorded during the trials. The presented environmental analyses did not identify measurable negative environmental effects associated with the introduction of the experience-based allocation concept. In particular, the results indicate that the concept is not expected to cause increased fuel consumption or additional CO_2 emissions due to operational inefficiencies. On the contrary, several analysed indicators even suggested operational performance comparable to, and in some cases slightly more efficient than, the reference scenarios where ATCOs worked in sectors with unit endorsement. Although the reasons for this effect cannot be conclusively determined based on the applied validation methodology, the recorded operational data suggest that ATCOs operating in unfamiliar sectors managed traffic in a particularly concentrated, careful and structured manner, thereby guiding aircraft along highly efficient routes in most analysed situations.

Overall, the presented findings indicate that the comparison between CBAS assessments and the experience-based allocation outcomes already provides encouraging results regarding the operational safety perspective of the concept, particularly due to the consistently low critical false rates observed across all evaluated configurations. At the same time, the high overall false rates demonstrate that the current UASLI-based allocation methodology still requires further refinement. Possible areas for future improvement include the selection and weighting of skill parameters, the aggregation logic, the interpretation of transferable competencies and the threshold definitions applied for allocation decisions.

Furthermore, the analyses showed the importance of complementing behavioural assessment methodologies with objective operational performance indicators. The first evaluations of simulation data already showed promising results, with regard to the absence of measurable negative environmental effects and the operationally efficient handling of traffic within unfamiliar sectors. These findings indicate that objective operational data may provide important additional insights beyond behavioural assessments alone and may support a more differentiated evaluation of transferable operational competencies.

At the same time, the presented results also underline that operational suitability cannot be assessed only on the basis of objective measurement data. Factors such as workload perception, confidence, uncertainty, stress and subjective operational comfort equally influence operational performance and decision-making. The experience-based allocation concept should therefore not be understood as a replacement for human judgement or existing operational concepts. Instead, the concept is intended to support and complement operational allocation processes by providing additional objective information for decision support.

6. Acknowledgement

The authors would like to thank all participating MUAC ATCOs and the entire MUAC team for their commitment, support and contributions throughout the preparation and execution of the validation campaign.

The IFAV3 project has received funding from the SESAR 3 Joint Undertaking under grant agreement No 101114683 under European Union’s Horizon Europe research and innovation programme.

7. Contact Author Email Address

mailto: jana.meier@dlr.de

8. Copyright Statement

The authors confirm that they, and/or their company or organisation, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.

References

[1] M. Baumgartner, ATCO Shortage: The Perfect Storm, IFATCA, 11th May 2025. [Online]. Available: https://ifatca.org/article/atco-shortage-the-perfect-storm/ [Accessed: 08th May 2026].

[2] A. G. Vicario, A. Levantesi, M. Finke, J. Meier and M. Terenzi, "Using Flexible Controller Validation Methods to Advance the Endorsement Practices in a Remote Tower Centre - A First Survey," 2025 Integrated Communications, Navigation and Surveillance Conference (ICNS), Brussels, Belgium, 2025, pp. 1-7, doi: 10.1109/ICNS65417.2025.10976920

[3] Commission Regulation (EU) 2015/340 of 20 February 2015: laying down technical requirements and administrative procedures relating to air traffic controllers' licences and certificates," Official Journal of the European Union, L 63, pp. 1- 122, 2015.

[4] SESAR Joint Undertaking, "IFAV3 - Increased Flexibility Across Verticals 3," 2023. [Online]. Available: https://www.sesarju.eu/projects/IFAV3. [Accessed: May 27, 2026].

[5] J. Meier, M. -M. Temme, R. Hunger and M. Finke, "Unlocking Potential: Enhancing Air Traffic Controller Deployment Through Dynamic Competency-Based Rostering," 2025 AIAA DATC/IEEE 44th Digital Avionics Systems Conference (DASC), Montreal, QC, Canada, 2025, pp. 1-10, doi: 10.1109/DASC66011.2025.11257372

[6] M. -M. Temme, J. Meier, R. Hunger, M. Finke, R. Abdellaoui, S. Shetty and A. D. Cicvara (2025). Development Development of technical and procedural requirements to support a system of flexible endorsement for upper airspace air traffic controllers. Deutscher Luft-und Raumfahrtkongress 2025 (DLRK 2025). Augsburg, Germany. DOI:10.25967/650088

[7] N. Brandenburger, I. Bier, M. Busse, J. Gärtner, T. Melzer, A technical demonstration of remote train operations using 5G mobile communications, SIGNAL+DRAHT, Vol. 115, No. 6, 2023.

[8] A. Simon, O. Gluchshenko and N. Wüstenbecker, "Transferring Air Traffic Arrival Management Concepts to the Maritime Domain: The Port of Hamburg Example," 2025 AIAA DATC/IEEE 44th Digital Avionics Systems Conference (DASC), Montreal, QC, Canada, 2025, pp. 01-10, doi: 10.1109/DASC66011.2025.11257181

[9] International Federation of Air Traffic Controllers' Associations (IFATCA), "Staff shortage survey – EUR Region," IFATCA, 2024. [Online]. Available: https://ifatca.org/staff-shortage-survey-eur-region/. [Accessed: May 27, 2026].

[10] SESAR Joint Undertaking, IFAV3 Solution 1 SPR-INTEROP / OSED (Final) Part I, Edition 1.0, to be published 31 January 2026.

[11] SESAR Joint Undertaking, SESAR Solution PJ.33-W3-01a SPR-INTEROP / OSED – Part I, Deliverable D2.1.001, Edition 02.01.02, 3 May 2023.

[12] SESAR Joint Undertaking, More flexible approach to air traffic controller endorsements, SESAR JU. [Online]. Available: https://www.sesarju.eu/news/more-flexible-approach-air-traffic-controller-endorsements/ [Accessed: 08th May 2026].

[13] R. Abdellaoui, I. Stasicka, M. Finke, A. Pick and N. Ahrenhold (2023). Supporting Air Traffic Controllers in handling sector specific tasks, enabled by the use of the Boundary Arrival Task Manager. IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC). Barcelona, Spain. DOI:10.1109/DASC58513.2023.10311241

[14] SESAR Joint Undertaking, "PJ33-W3 FALCO - Flexible ATCO Endorsement and LDACS Complement," SESAR Projects Portal. [Online]. Available: https://www.sesarju.eu/projects/FALCO. [Accessed: May 27, 2026].

[15] M. Finke, R. Hunger, L. Tyburzy, and J. Meier, "Using tool support to enable a more flexible deployment of air traffic controllers to sectors", in Proc. Deutscher Luft- und Raumfahrtkongress (DLRK 2024), Hamburg, Germany, Sep. 30 – Oct. 2, 2024. doi: 10.25967/630243

[16] SESAR Joint Undertaking, "IFAV3 - Increased Flexibility Across Verticals 3," 2023. [Online]. Available: https://www.sesarju.eu/projects/IFAV3. [Accessed: May 27, 2026].

[17] J. Meier, M. -M. Temme, R. Hunger and M. Finke (2025). Upper Airspace Skill-Level Index for Smart Competency Monitoring. Projektbericht. DLR-IB-FL-BS-2024-71. 47 S.

[18] J. Meier, M. -M. Temme, R. Hunger, M. Finke (2025). Upper Airspace Skill-Level Index for Minimum Competency Level Prediction. DLR-Interner Bericht. DLR-IB-FL-BS-2025-43. 38 S.

[19] Oprins, E., Burggraaff, E., & van Weerdenburg, H. (2006). Design of a Competence-Based Assessment System for Air Traffic Control Training. The International Journal of Aviation Psychology, 16(3), 297–320. https://doi.org/10.1207/s15327108ijap1603_4