ISSN 2669-2023 (print) / ISSN 2538-8959 (on-line) / **https://doi.org/10.47459/cndcgs.2026.135**

Modeling and Demonstrating Air Traffic Controller Workload Using Simulation in Academic Training

Jan MAZANEK[1]*, Tomáš PECHACEK[1], Vladimir SMRZ[1], Jiri ULVR[1]

[1] Air Force Department, Faculty of Military Technology, University of Defence, Kounicova 65, 602 00, Brno, Czech Republic

*Correspondence: * jan.mazanek@unob.cz*

Abstract

This study explores whether simulator-derived workload metrics can indicate differences in performance between ATC student cohorts. A standardized radar simulation was conducted with 14 participants (7 fourth-year, 7 fifth-year). Workload was modeled using event-based HUM-SIM principles and combined with instructor evaluation. The results suggest higher workload exposure, longer critical states, and greater deviation from the planned scenario in less experienced students. Differences appear not only in workload magnitude but in its temporal regulation. These findings indicate that workload-informed metrics may support objective performance assessment in simulator-based training, although further validation on larger datasets is required.

KEY WORDS: *Air Traffic Control; Workload; Simulation-based Training; Human Factors; Academic ATC Education; Human Performance*

Citation: Mazánek, J.; Pecháček, T.; Smrž, V.; Ulvr, J. Modeling and Demonstrating Air Traffic Controller Workload Using Simulation in Academic Training. In Proceedings of the Challenges to National Defence in Contemporary Geopolitical Situation, Brno, Czech Republic, 2026. ISSN 2538-8959. https://doi.org/10.47459/cndcgs.2026.135

1. Introduction

Air Traffic Controller (ATCO) workload represents one of the most extensively examined constructs in aviation human factors research. Within Air Traffic Management (ATM) systems, controller performance is directly associated with operational safety, traffic efficiency, and overall system resilience under dynamic conditions [1; 2]. Due to the time-critical, information-dense and safety-sensitive nature of Air Traffic Control (ATC) operations, inappropriate workload levels may influence situational awareness, decision-making processes and error management strategies [1; 3]. Consequently, understanding, modeling and managing ATCO workload remains an important topic within aviation psychology and safety-oriented research.

Mental workload in ATC is commonly described as a multidimensional construct emerging from the interaction between task demands, environmental complexity, technological systems and the cognitive resources of the operator [2; 4]. Workload-generating elements typically include conflict detection and resolution, traffic sequencing and vectoring, communication management, coordination tasks and handling of abnormal and emergency situations [4]. These operational demands are dynamically interrelated and closely connected to situational awareness and Threat and Error Management (TEM) principles [5].

Various methodological approaches have been developed to assess ATCO workload. Subjective tools such as the NASA Task Load Index (NASA-TLX) are widely used due to their multidimensional structure and practical applicability [6]. However, subjective techniques reflect perceived workload retrospectively and may be influenced by individual interpretation [7]. Complementary approaches, including performance-based indicators and physiological measurements, have therefore been explored in the literature [1; 2]. Recent methodological overviews emphasize the importance of combining different modalities depending on the specific research objective and context [8]. Despite extensive research efforts, no universally accepted single metric of ATCO workload currently exists.

Simulation has become a central instrument for both ATCO training and workload research. ICAO and EUROCONTROL documents identify simulation-based training as a fundamental component of competency-based training and assessment systems [9; 10]. Simulation environments allow controlled manipulation of traffic density, conflict structure and task complexity, thereby enabling repeatable workload studies under standardized conditions [8]. While most empirical investigations have been conducted in certified training organizations or operational Air Navigation Service Provider (ANSP) environments, academic institutions offering ATC education may also provide a suitable platform for exploratory workload modeling.

In the context of military ATCO education, the modernization of selection and training processes has recently been discussed, particularly in relation to generational changes and the integration of simulation technologies [11]. Particular attention has been paid to the development of alternative training models that better reflect operational requirements and evolving training needs [12]. These considerations are further supported by findings indicating changing characteristics of Generation Z applicants, including differences in theoretical preparedness and emerging trends in health status, which may influence suitability for demanding aviation specializations [13]. At the same time, recent research highlights the need to systematically address praxeological aspects of professional training systems to ensure effective preparation of aviation personnel [14]. The integration of simulation-based elements into selection and training has been proposed as a way to better align assessment methods with contemporary educational approaches and evolving candidate profiles [11].

Building upon these considerations, the present study examines whether a structured, performance-based workload model derived from simulator data can reflect differences between student cohorts with varying levels of experience. The objective is not operational validation but rather to explore the potential of workload modeling as a supportive analytical tool within an academic ATCO training environment.

2. Methods

The simulated exercise implemented in this study was based on the EUROCONTROL Training Plan – Module 6 (Approach Surveillance Rating) [15]. The scenario (see Fig. 1) represents a standardized radar-based approach control exercise designed to assess student performance in the role of Approach Controller (APP).

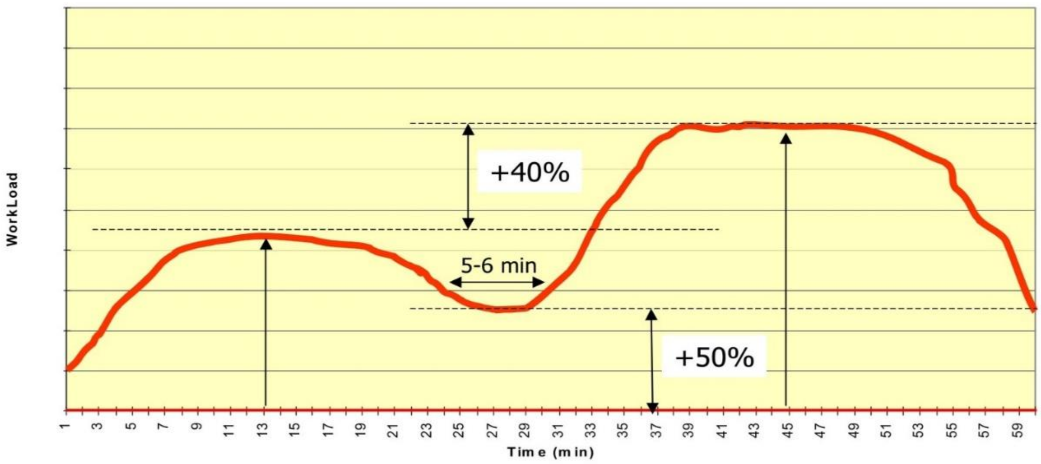

Fig. 1 Planned workload profile according to Training Plan – Module 6 *[16]*.

The original scenario duration (45–50 minutes) was reduced to 30 minutes while preserving the relative workload structure and temporal distribution of task demands in accordance with the training plan specifications (see Fig. 2) [16]. The workload profile follows the EUROCONTROL design principles with predefined peak–valley dynamics reflecting different phases of task complexity.

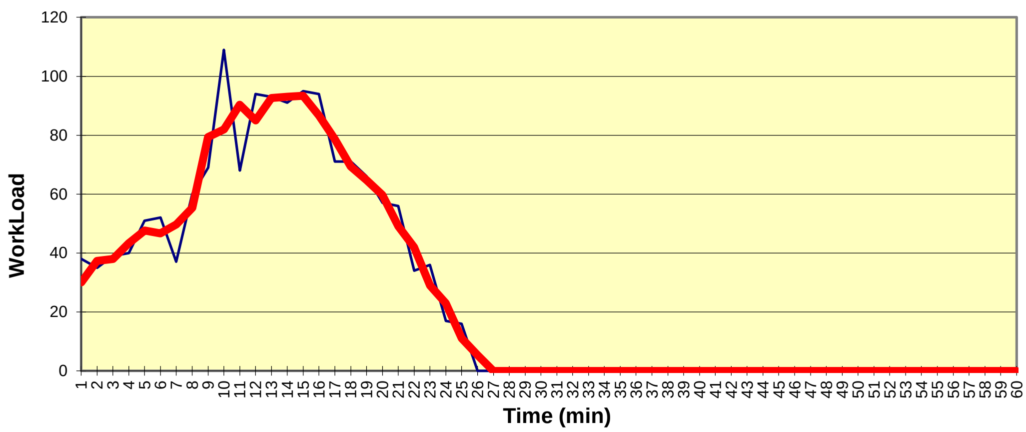

Fig. 2 Designed workload profile of the experimental scenario.

The scenario was conducted in Class D airspace within the vertical range GND to FL095 and included a controlled aerodrome with standard arrival (STAR) and departure (SID) procedures, including holding patterns. The traffic structure comprised a mix of arrivals and departures and included conflict situations, ensuring realistic variability in operational workload under standard operational conditions. The scenario design was kept constant across all participants to ensure comparability of workload dynamics and performance outcomes.

2.1. Workload Model

Workload in this study is defined as a quantitative indicator representing the aggregated sum of operational actions, events, or states occurring within a defined time window during the simulation. Each event reflects physical and cognitive demand imposed on the Air Traffic Controller (ATCO) and allows comparison of workload intensity across different temporal segments of the same exercise.

Each workload factor (see Table 1) is considered independent and contributes a predefined relative weight ("points") to the total workload. The absolute value of individual weights is not critical; rather, the proportional relationship between factors expresses their relative operational complexity. The selection of factors and their assigned weights follows EUROCONTROL HUM-SIM principles and depends on course type, training phase, and local instructional objectives [16].

Table 1.

Examples of workload factor weighting

Problem Identifier	Short Description	Points
UNKN	Monitoring an unknown acft	5
COO_S	Standard Coordination for the unit	4
C_60	Traffic on 60 deg tracks in conflict	8
VEC	While acft on vectors …	10
ARRVS	Arrival on vectors and sequencing	19
EMERGENCY	For aircraft in danger	20

For each time window Δt (typically 1 minute), workload is calculated as:

$$WL = \frac{\sum p_j}{\Delta t} + N \cdot \frac{MAXLOAD}{MAX\_ACFT} \qquad (1)$$

where: p_j= weight of factor $j$active in the given minute; $\Delta t = 1$min; N= number of monitored aircraft; MAX_ACFT= aircraft count corresponding to radio frequency saturation; $MAXLOAD$= maximum additional workload attributed to frequency occupancy.

Radio frequency occupancy is considered a significant workload contributor in ATC, as sector configuration and traffic capacity are often constrained by communication load [11]. The threshold for "radio saturation" is determined by scenario design judgment. According to EUROCONTROL guidance:

$$MAX\_ACFT = 25;$$
$$MAXLOAD = 300.$$

The additional workload due to frequency occupancy is therefore computed linearly as:

$$extra\ points = N \cdot \frac{MAXLOAD}{MAX\_ACFT} \qquad (2)$$

The resulting minute-by-minute workload time series forms the Workload Graph. To reflect cognitive inertia and avoid unrealistic discontinuities, Laplacian smoothing is applied to the raw workload curve in accordance with EUROCONTROL recommendations [16]. The smoothed curve is considered the final workload profile (see Fig. 2).

2.2. Participants and Data Collection

The study involved 15 undergraduate students enrolled in the Military Air Traffic Control specialization at the University of Defence, Brno. Due to technical issues, one dataset was excluded, resulting in a final sample of 14 participants (7 fourth-year and 7 fifth-year students).

Only senior students were included, as the execution of the standardized radar scenario required prior training in radar control procedures and operational decision-making.

Participants were familiarized with the simulator environment prior to testing. Fifth-year students had prior experience, while fourth-year students received additional familiarization to ensure basic operational competence.

Each participant completed one simulation run, and performance data were recorded under anonymized identifiers and exported for subsequent workload computation and statistical analysis.

2.3. Simulation Environment

The experiment was conducted using the CASS-UNOB air traffic control simulator, version G4.1.X (ARTISYS Ltd.), which supports radar-based ATC training under controlled and repeatable conditions (see Fig. 3).

Each participant operated from a standard radar controller workstation. Traffic behavior was controlled by a pseudo-pilot to ensure consistency across all experimental runs.

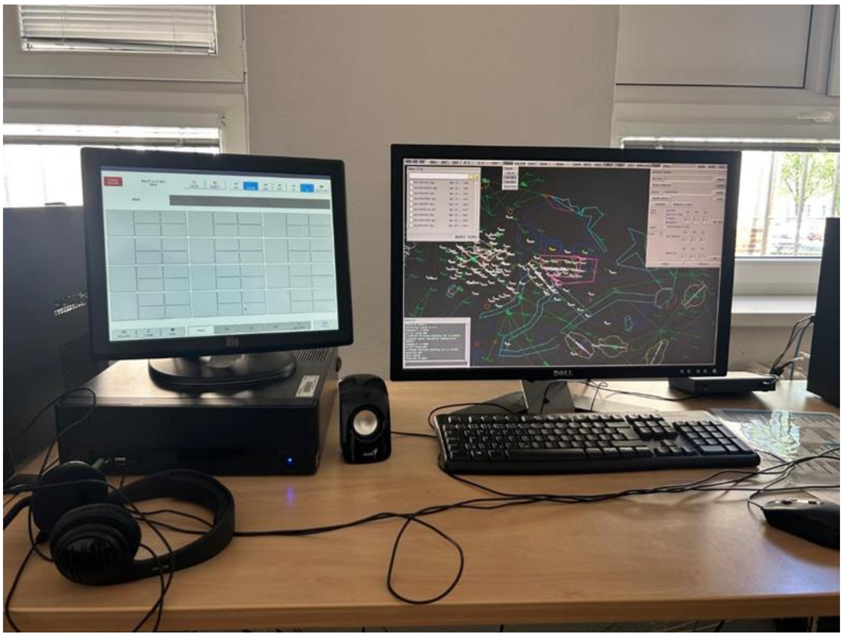

Fig. 3 Experimental workstation configuration.

2.4. Scenario Administration and Data Recording

A single standardized scenario (see Fig. 2) with a predefined and ideally structured workload profile, designated as "Baseline," was used for all participants. The scenario was designed by instructors at the University of Defence to evaluate whether the simulator environment could reliably induce the intended workload structure and to observe participant responses under comparable conditions.

Participants were not informed about the specific content or timing of events within the scenario prior to testing. Before the experimental session, all students were familiarized with the simulator interface and operational functions. This familiarization phase was particularly important for fourth-year students, who had limited prior exposure to this specific simulation configuration. Fifth-year students had previous experience; however, they were also provided with the opportunity to clarify procedural uncertainties before testing.

Each participant's performance data were recorded under a unique anonymized identifier:

1. Subject02–Subject08 (fourth-year students);
1. Subject09 and higher (fifth-year students).

The duration of individual simulation runs ranged from 0 to 30 minutes, depending on traffic handling efficiency and scenario progression. Upon completion of each session, simulator logs were stored and exported for subsequent workload computation and statistical analysis.

The experimental setup was designed to isolate workload-related effects while minimizing variability in scenario structure.

2.5. Integrated Assessment Model of Simulator Exercise Performance and Statistical Framework

To enable transparent, repeatable, and analytically defendable evaluation of a standardized ATC simulator exercise, an integrated assessment framework was defined that combines:

1. **Objective workload indicators** computed from simulator logs using EUROCONTROL HUM-SIM principles and the scenario baseline profile.
2. **Formative instructor competency ratings** collected using a standardized CBTA-aligned assessment form.

The framework is intended for **analytical comparison in an academic pilot setting** rather than operational validation or certification.

The study addresses three linked objectives:

1. Objective 1 (primary): Cohort differentiation.

Determine whether simulator-derived workload indicators differ between student cohorts with different experience levels (4th-year vs 5th-year).

2. Objective 2 (secondary): Workload–performance association.
Examine whether workload exposure is statistically associated with instructor-rated procedural performance.
3. Objective 3 (descriptive): Baseline conformity and temporal structure.
Quantify how measured workload evolution corresponds to the planned baseline profile and whether workload maxima shift in time.

Data structure and workload time series. For each participant $i = 1, \dots, n$, the simulator log data were used to compute a one-minute workload time series:

$$WL_i(t), t = 1, \dots, T \tag{3}$$

where $T = 30$minutes (scenario duration). A designed scenario baseline workload profile was defined as:

$$WL_{base}(t), t = 1, \dots, T \tag{4}$$

The resulting dataset therefore consists of:

1. a workload matrix $\mathbf{WL} \in \mathbb{R}^{n \times T}$;
2. individual time-series trajectories (participant-level curves);
3. a baseline reference trajectory (designed workload profile).

This structure enables both time-domain comparisons (temporal evolution) and summary indicator computation (participant-level workload descriptors).

Derived workload indicators. From each participant's workload trajectory $WL_i(t)$, a set of workload indicators was derived to represent overall exposure, peak burden, instability, and conformity to the designed baseline.

Global exposure measures:

Mean workload

$$\bar{W}L_i = \frac{1}{T}\sum_{t=1}^{T} W\,L_i(t) \tag{5}$$

Maximum workload

$$WL_i^{max} = \max_t WL_i(t) \tag{6}$$

These two indicators summarize general workload exposure and peak intensity.

Workload zone exposure and critical workload episodes. To capture time spent under elevated or critical demand, workload zones were defined as:

1. **Normal zone:** $WL_i(t) \leq 99$.
2. **Elevated zone:** $99 < WL_i(t) \leq 109$.
3. **Critical zone:** $WL_i(t) > 109$.

The proportion of time spent in each zone was computed as:

$$z_{N,i} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{I}(W\,L_i(t) \leq 99);\ z_{E,i} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{I}(99 < W\,L_i(t) \leq 109);\ z_{C,i} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{I}(W\,L_i(t) > 109) \tag{7}$$

The zone thresholds (99 and 109 workload points) were scenario-referenced values derived from calibration of the designed baseline profile and used consistently across all participants to enable direct comparability within the same standardized exercise.

In addition, **critical workload episodes** were defined as contiguous time intervals satisfying $WL_i(t) > 109$. For each participant, two episode indicators were computed:

1. E_i^{count}: number of critical episodes,
2. E_i^{dur}: cumulative duration of critical episodes (minutes).

These indicators capture whether critical workload occurs as short spikes or sustained overload periods.

Workload volatility (temporal instability)

To quantify workload fluctuation dynamics, first differences were computed as:

$$\Delta WL_i(t) = WL_i(t) - WL_i(t-1), t = 2, \dots, T \tag{8}$$

Workload volatility was defined as the standard deviation of first differences:

$$Vol_i = SD(\Delta WL_i(t)) \tag{9}$$

Higher volatility values indicate stronger workload oscillations and reduced temporal stability.

Deviation from baseline workload profile

Scenario conformity was quantified by the mean absolute deviation from the baseline profile:

$$MAD_i = \frac{1}{T}\sum_{t=1}^{T} | WL_i(t) - WL_{base}(t) | \tag{10}$$

This metric captures the magnitude of divergence between observed and planned workload structure over time.

Peak timing (temporal workload shift)

To capture systematic temporal shifts of workload maxima, the time location of maximum workload was computed as:

$$t_i^{max} = arg\,\max_t WL_i(t) \tag{11}$$

Peak timing is reported descriptively and compared across cohorts to evaluate whether workload maxima occur in the intended scenario phase or shift due to cumulative task demands.

Composite Workload Index (WI)

To enable aggregation of heterogeneous workload indicators without scale distortion, selected indicators were standardized using z-score normalization:

$$Z_{k,i} = \frac{m_{k,i} - \mu_k}{\sigma_k} \tag{12}$$

where $m_{k,i}$ is the value of indicator k for participant i, and μ_k, σ_k are the sample mean and standard deviation for indicator k.

A reduced subset of workload components was retained to avoid over-parameterization in a pilot sample:

1. critical zone exposure $z_{C,i}$;
2. critical episode duration E_i^{dur} (or episode count E_i^{count}, depending on stability);
3. volatility Vol_i;
4. baseline deviation MAD_i.

The composite Workload Index was then defined as an equally weighted mean of standardized components:

$$WI_i = \frac{1}{K}\sum_{k=1}^{K} Z_{k,i} \tag{13}$$

Interpretation:

1. $WI_i = 0$ indicates average workload burden within the observed sample,
2. $WI_i > 0$ indicates above-average workload exposure and instability,
3. $WI_i < 0$ indicates below-average workload exposure.

Because WI is standardized, it is interpreted as a relative workload measure within the sample rather than an externally calibrated workload scale.

Instructor competency assessment and Competency Index (CI)

Instructor evaluation was conducted by a **single ATC instructor** using a **standardized formative assessment form applied identically to all participants**. The assessment comprised $L = 10$ competency criteria representing procedural correctness and operational discipline consistent with competency-based assessment practice.

Each criterion was graded on an ordinal scale:

$$\{A, B, C, D, X\}$$

where X denotes “not assessed”.

For quantitative analysis, grades were mapped to points:

$$g(A) = 4, g(B) = 3, g(C) = 2, g(D) = 1$$

Items marked X were excluded. Let $\mathcal{L}_i \subseteq \{1, \dots, L\}$ denote assessed items for participant i. The Competency Index was defined as:

$$CI_i = 100 \cdot \frac{\sum_{\ell \in \mathcal{L}_i} g_{\ell,i}}{4 \cdot | \mathcal{L}_i |} \tag{14}$$

Thus $CI_i \in [0,100]$represents a normalized instructor-rated competency score.

If any safety-critical criterion (for example separation minima violation, incorrect conflict resolution, or procedural non-compliance) was graded D, the participant was classified as unsatisfactory regardless of the composite CI. This rule ensures that severe safety deficiencies cannot be masked by high scores in non-critical criteria.

Integration of workload and competency: Overall Performance Index (OPI)

To explore the interaction between workload exposure and procedural performance, an Overall Performance Index was defined as a composite of CI and WI:

$$OPI_i = CI_i - \lambda WI_i \tag{15}$$

Because CI is scaled in $[0, 100]$and $WI$is a standardized index with typical variation within approximately ± 2SD for most observations, the scaling parameter $\lambda$was selected to provide a meaningful, non-dominant workload influence on the composite score.

In this study, $\lambda$was set so that a one-standard-deviation increase in workload corresponds to a 10-point change in the composite performance score:

$$\lambda = 10$$

The final definition therefore becomes:

$$OPI_i = max(0, CI_i - 10\, WI_i) \tag{16}$$

Interpretation:

1. if $WI_i = 0$, then $OPI_i = CI_i$;
2. a +1 SD increase in workload decreases OPI by 10 points;
3. a −1 SD deviation increases OPI by 10 points.

Workload is not interpreted as inherently detrimental. The composite index is used to identify cases where elevated workload burden co-occurs with reduced procedural stability in instructor assessment.

Statistical analysis plan

All statistical analyses were performed at:

$$\alpha = 0.05$$

Given the pilot sample size, both statistical significance and effect sizes are reported and interpreted cautiously.

Cohort comparison (primary objective). To test whether workload exposure differs between cohorts (4th-year vs 5th-year), group comparisons were performed for:

1. $\bar{WL}, WL^{max}$;
2. z_E, z_C, episode indicators;
3. Vol, MAD;
4. composite WI.

Normality was evaluated using the Shapiro–Wilk test. If assumptions were satisfied, independent-samples t-tests were applied. Otherwise, the Mann–Whitney U test was used. Effect sizes were reported as Cohen's d(parametric) or rank-biserial correlation (non-parametric).

Baseline conformity and temporal structure (baseline objective). Baseline conformity was quantified using $MAD$and descriptive comparison of time-series structure. Peak timing $t^{max}$was reported and compared between cohorts to identify systematic temporal shifts of workload maxima relative to the designed baseline scenario.

Workload–performance association (secondary objective). Associations between workload exposure and instructor-rated competency were examined using correlation and regression:

1. Pearson correlation if normality assumptions were satisfied, otherwise Spearman's rank correlation.
2. Linear regression model:

$$CI_i = \beta_0 + \beta_1 WI_i + \varepsilon_i \tag{17}$$

Because WI is standardized, $\beta_1$represents the expected change in CI per one SD increase in workload exposure. The analysis is exploratory and does not imply causality.

Methodological scope and interpretation. This integrated assessment framework is intended for exploratory analysis of workload dynamics and performance differences in an academic simulator training setting. The limited sample size and single-scenario design restrict external generalizability, and the derived indices should be interpreted as comparative measures within the observed dataset. The framework provides a methodological basis for subsequent studies with larger samples, multiple scenarios, and expanded measurement modalities.

3. Resalts

Descriptive statistics and inferential results are reported without interpretation and their implications are addressed in the Discussion section. The final dataset comprised 14 participants (7 fourth-year and 7 fifth-year students). Group-level comparisons across workload and performance indicators are summarized in Table 2.

Table 2.

Cohort comparison of workload and performance indicators

Metric	n4	n5	Mean4	Mean5	p	Test	Stat	Effect	EffectName
meanWL	7	7	72,590	57,737	0,001	ttest2	4,190	2,239	Cohen_d
maxWL	7	7	117,514	100,400	0,024	ttest2	2,592	1,386	Cohen_d
zE	7	7	0,097	0,041	0,140	ttest2	1,580	0,845	Cohen_d
zC	7	7	0,124	0,005	0,037	ranksum	68,500	-0,653	rank_biserial
E_count	7	7	1,000	0,143	0,070	ranksum	67,500	-0,612	rank_biserial
E_dur	7	7	3,857	0,143	0,037	ranksum	68,500	-0,653	rank_biserial
Vol	7	7	11,890	10,879	0,230	ttest2	1,265	0,676	Cohen_d
MAD	7	7	25,733	14,018	0,004	ttest2	3,604	1,926	Cohen_d
WI	7	7	0,518	-0,518	0,025	ttest2	2,569	1,373	Cohen_d
CI	7	7	72,619	83,730	0,056	ttest2	-2,119	-1,132	Cohen_d
OPI	7	7	67,442	88,907	0,013	ttest2	-2,921	-1,561	Cohen_d

Across the primary workload indicators, the fourth-year cohort exhibited higher values than the fifth-year cohort. Mean workload was higher in fourth-year students (72.59 vs. 57.74, $p = 0.0013$, $d = 2.24$). Maximum workload also reached higher values in the fourth-year group (117.51 vs. 100.40, $p = 0.0236$, $d = 1.39$).

Differences were observed in the critical workload zone (zC). Fourth-year students spent a greater proportion of time in critical workload conditions than fifth-year students (0.124 vs. 0.0046, $p = 0.0373$, rank-biserial = −0.653). The duration of critical episodes (E_dur) was longer in the fourth-year cohort (3.86 min vs. 0.14 min, $p = 0.0373$).

The number of critical episodes (E_count) was higher in fourth-year students (1.00 vs. 0.14), although this difference did not reach statistical significance ($p = 0.0699$). Time spent in the elevated workload zone (zE) did not differ significantly between groups ($p = 0.140$).

Workload volatility (Vol) showed no statistically significant difference between cohorts ($p = 0.230$).

Baseline conformity, expressed as mean absolute deviation (MAD), differed between groups. Fourth-year students exhibited higher deviation from the planned workload profile (25.73 vs. 14.02, $p = 0.0036$, $d = 1.93$).

The composite Workload Index (WI) was higher in fourth-year students (0.52 vs. −0.52, $p = 0.0246$, $d = 1.37$).

Instructor-rated performance (CI) was higher in the fifth-year cohort (72.62 vs. 83.73), with the difference approaching statistical significance ($p = 0.0557$). The Overall Performance Index (OPI) differed significantly between cohorts (67.44 vs. 88.91, $p = 0.0128$, $d = 1.56$).

Table 3 presents participant-level values for selected indicators, including WI, CI, and OPI.

Table 3.

Individual participant metrics (WI, CI, OPI and components)

Participant	meanWL	maxWL	zN	zE	zC	E_count	E_dur	Vol	MAD	tmax	WI	Z_zC	Z_E_dur	Z_Vol	Z_MAD	CI	OPI
Subject02	72,2	114,6	0,7	0,2	0,1	1,0	3,0	9,8	24,3	12,0	0,0	0,3	0,3	-1,0	0,5	72,2	72,1
Subject03	80,0	141,8	0,6	0,0	0,4	2,0	12,0	13,9	33,4	12,0	2,2	2,8	2,8	1,7	1,6	63,9	41,6
Subject04	64,6	104,2	0,8	0,2	0,0	0,0	0,0	10,9	17,8	15,0	-0,4	-0,6	-0,6	-0,3	-0,2	80,6	84,8
Subject05	61,2	98,8	1,0	0,0	0,0	0,0	0,0	10,6	14,5	19,0	-0,6	-0,6	-0,6	-0,5	-0,6	69,4	75,2
Subject06	77,3	138,6	0,7	0,0	0,2	1,0	7,0	13,1	30,5	20,0	1,3	1,4	1,4	1,1	1,3	83,3	70,3
Subject07	76,3	110,0	0,8	0,1	0,0	1,0	1,0	10,6	29,7	21,0	0,0	-0,3	-0,3	-0,5	1,2	66,7	66,4
Subject08	76,5	114,6	0,7	0,2	0,1	2,0	4,0	14,2	30,1	11,0	1,1	0,6	0,6	1,9	1,2	72,2	61,7
Subject09	56,4	100,8	1,0	0,0	0,0	0,0	0,0	10,2	12,1	12,0	-0,7	-0,6	-0,6	-0,8	-0,9	88,9	96,0
Subject10	54,3	100,6	1,0	0,0	0,0	0,0	0,0	12,0	10,1	16,0	-0,5	-0,6	-0,6	0,4	-1,2	94,4	99,2
Subject15	65,3	97,8	1,0	0,0	0,0	0,0	0,0	9,6	18,9	16,0	-0,6	-0,6	-0,6	-1,2	-0,1	86,1	92,2
Subject16	63,7	104,0	0,9	0,1	0,0	0,0	0,0	12,4	17,8	17,0	-0,2	-0,6	-0,6	0,7	-0,2	69,4	71,3
Subject17	51,0	95,4	1,0	0,0	0,0	0,0	0,0	10,6	8,2	11,0	-0,7	-0,6	-0,6	-0,5	-1,4	77,8	85,3
Subject18	62,8	110,6	0,8	0,2	0,0	1,0	1,0	11,4	20,1	19,0	-0,1	-0,3	-0,3	0,0	0,0	69,4	70,7
Subject19	50,6	93,6	1,0	0,0	0,0	0,0	0,0	10,0	10,9	15,0	-0,8	-0,6	-0,6	-0,9	-1,1	100,0	107,8

The individual results show variability within both cohorts. Several fourth-year participants exhibited positive WI values, while most fifth-year participants showed negative WI values.

No participant met the fail criterion, and no case was classified as unsatisfactory. The temporal evolution of workload for all participants is shown in Fig. 4.

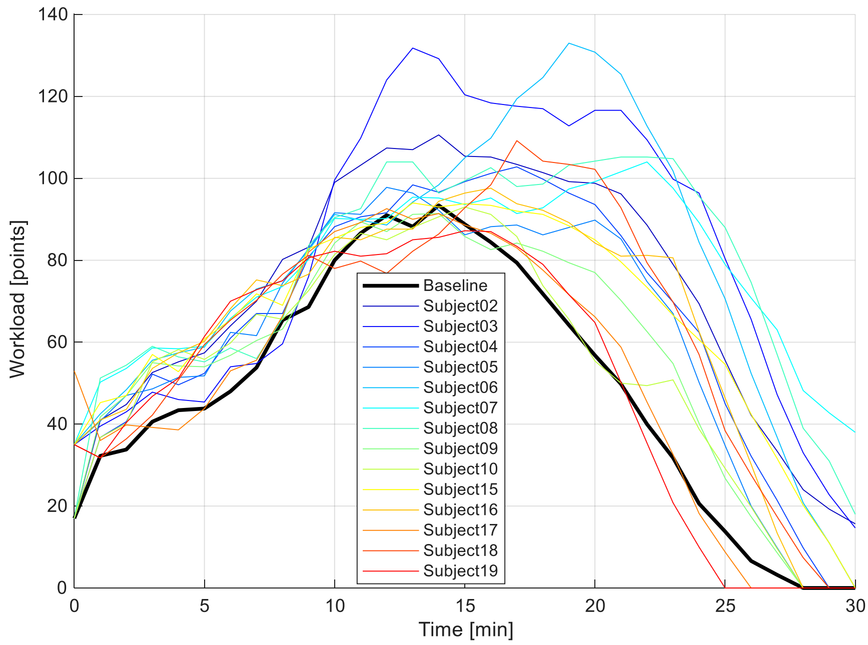

Fig. 4 Workload time series for all participants and baseline profile.

Both cohorts reached peak workload at approximately the same time (around minute 15). Differences are observable in the post-peak phase (minutes 18–25), where higher workload levels persist in the fourth-year cohort (Fig. 4).

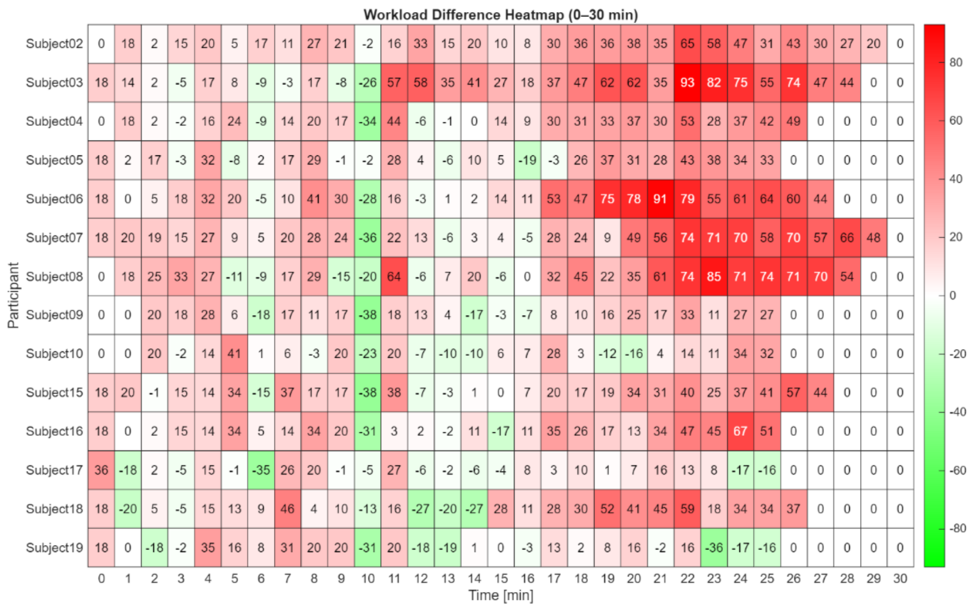

Fig. 5 Heatmap of workload deviation from baseline.

Figure 5 presents workload deviations relative to the baseline profile. Positive deviations occur predominantly in the later phase of the scenario (approximately minutes 20–25), while smaller deviations are observed in the fifth-year cohort.

4. Discussion

The results demonstrate systematic differences between student cohorts with different levels of experience. The less experienced cohort exhibited higher workload exposure, more frequent and prolonged critical workload states, and greater deviation from the intended scenario structure. These differences are consistently reflected across multiple independent indicators, including mean workload, critical workload exposure, baseline deviation (MAD), and the composite Workload Index.

The observed differences are not limited to workload magnitude. Both cohorts reached peak workload at a comparable time; however, the post-peak workload evolution differs. The more experienced cohort shows a faster reduction of workload following peak traffic complexity, whereas the less experienced cohort maintains elevated workload levels into later phases of the scenario. This pattern reflects differences in the ability to regulate workload over time under dynamic conditions and is consistent with findings in ATC human factors research, where workload is closely linked to dynamic task management and operator capacity limits [1; 2].

This behaviour is consistent with established models of expertise in complex systems, where experienced operators demonstrate more effective task prioritization, anticipatory control, and resolution of emerging conflicts. In the context of air traffic control, this corresponds to differences in situational awareness and temporal management of task demand as described in prior studies [1; 4].

A negative association between workload exposure and instructor-rated competency is observed across the dataset. Although the regression model does not reach statistical significance, the direction and consistency of the relationship indicate that increased workload burden is associated with reduced procedural performance within this sample. This relationship is aligned with theoretical models of mental workload, where performance degradation occurs when task demands exceed available cognitive resources [2; 4].

The introduction of the Overall Performance Index (OPI) provides an integrated view of performance under workload conditions. The results show that comparable competency scores (CI) can correspond to different workload contexts. Incorporating workload into performance assessment therefore enables differentiation between performance achieved under lower versus higher operational strain, which corresponds to recommendations emphasizing multimodal workload assessment approaches [8].

Differences in conformity to the baseline workload profile further distinguish the cohorts. The more experienced cohort follows the intended temporal structure more closely, while the less experienced cohort shows larger deviations, particularly in the later phase of the scenario. This reflects the role of structured workload design and temporal workload distribution in simulation-based training as defined in EUROCONTROL methodologies [16]. The workload model generates multiple complementary indicators capturing workload magnitude, critical exposure, temporal instability, and baseline conformity. The composite Workload Index (WI) provides a compact representation of these dimensions; however, the equal weighting of components represents a simplifying assumption and should be interpreted accordingly.

The integration of workload and competency into a combined metric (OPI) represents a structured approach to linking objective simulator data with instructor-based evaluation. While the current formulation is exploratory, it establishes a basis for further methodological development and validation.

Several limitations constrain the interpretation of the results. The small sample size limits statistical power and generalizability. The use of a single standardized scenario restricts applicability across different operational contexts. The workload model represents a simplified abstraction based on predefined event weights and does not incorporate subjective or physiological measures. Instructor evaluation was conducted by a single rater, and inter-rater reliability was not assessed. The composite indices are based on within-sample normalization, which limits comparability across datasets.

Future research should extend the proposed framework using larger samples, multiple scenarios, and additional measurement modalities. Further work is required to evaluate alternative weighting schemes and to validate the predictive value of the proposed indices. From an applied perspective, the integration of workload-based metrics into simulator training offers a pathway toward more objective and data-driven performance assessment.

5. Conclusions

The results demonstrate that simulator-derived workload indicators differentiate between student cohorts with different levels of experience in an academic ATC training environment. Less experienced students exhibited higher workload exposure, more frequent critical workload conditions, and greater deviation from the intended scenario structure.

The primary difference was not limited to workload magnitude but was reflected in the temporal regulation of workload, particularly in the post-peak phase of the exercise. The integration of workload metrics with instructor-based assessment through a composite performance index further revealed differences in performance under varying workload conditions.

Although the study is based on a pilot sample, the findings support the use of workload modeling as a complementary tool for objective performance analysis in simulator-based training. The proposed framework provides a structured basis for integrating workload dynamics into competency-based assessment and can support further development of data-driven evaluation approaches in ATC training.

Acknowledgements. This work was supported by the institutional support of the Ministry of Defence of the Czech Republic.

References

1. **Wickens C.D., Helton W.S., Hollands J.G., Banbury S.** Engineering psychology and human performance. Fifth edition. New York: Routledge, 2021. ISBN 9781032011738.

2. **Young M.S., Brookhuis K.A., Wickens C.D., Hancock P.A.** State of science: mental workload in ergonomics. Ergonomics. Roč. 58, č. 1, pp. 1-17. ISSN 0014-0139.
3. **Stein E.S.** Air traffic controller workload: An examination of workload probe. New Jersey: United States. Department of Transportation. Federal Aviation Administration, 1985.
4. **Loft S., Sanderson P., Neal A., Mooij M.** Modeling and Predicting Mental Workload in En Route Air Traffic Control: Critical Review and Broader Implications. Human Factors. Roč. 49, č. 3, pp. 376-399. ISSN 0018-7208.
5. **ICAO.** Doc 9683: Human Factors Training Manual. Online. In: International Civil Aviation Organization. Montreal, 1998, p. 451.
6. **Hart S.G., Staveland L.E.** Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research: Results of empirical and theoretical research. In: Advances in psychology. 52. Elsevier, 1988, p. 139-183. ISSN 0166-4115.
7. **Rubio S., Díaz E., Martín J., Puente J.M.** Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. Applied Psychology. Roč. 53, č. 1, pp. 61-86. ISSN 0269-994X.
8. **Ulvr J., Mazánek J., Pecháček T., Smrž V., Engelová P.** Theoretical Background and Methodological Approaches for the Assessment of the Professional Workload of air Traffic Controllers. In: 2025 New Trends in Aviation Development (NTAD). 2025, p. 315-322. ISSN 2836-2764.
9. **ICAO.** Doc 9868: Procedures for Air Navigation Services – Training. 2nd Edition. ICAO, 2015. ISBN 978-92-9249-874-0.
10. **ICAO.** Doc 10056: Manual on Air Traffic Controller Competency-based Training and Assessment. Online. In: International Civil Aviation Organization. Montreal, 2016, p. 451.
11. **Ulvr, J., Pecháček, T., Mach O., Klímová T., Račková P.** Quality Improving of the Air Traffic Controller Selection Process at the University of Defence. Advances in Military Technology. Roč. 19, č. 2, pp. 301-315. ISSN 2533-4123.
12. **Jevčák J., Rozenberg R., Kal'avský P., Kelemen M., Klír R. et al.** Identification of Praxeological Problem of Professional Training of Personnel for a Specific Air Operator. In: 2020 New Trends in Aviation Development (NTAD). 2020, p. 119-122.
13. **Ulvr J., Račková P., Smrž V., Pecháček T., Mach O.** Life Expectancy of Generation Z versus Requirements for the Military Pilot and Air Traffic Controller Study Specializations at the University of Defence. Online. In: 2023 21st International Conference on Emerging eLearning Technologies and Applications (ICETA). Stary Smokovec, Slovakia: IEEE, 2023, p. 542-550. ISBN 979-8-3503-7069-0.
14. **Harčár I., Antoško M., Fábry Ň., Némethová H., Šuťák Ň. et al.** Training Models for Military Air Traffic Controllers. In: 2020 New Trends in Aviation Development (NTAD). 2020, p. 89-92.
15. **EUROCONTROL.** Approach Control Surveillance Rating with Radar and Terminal Endorsements: Training Plans. Online. 2nd Edition. Luxembourg, 2003.
16. **EUROCONTROL.** HUM-SIM: Principles for the Design of Radar Simulation Exercises and Course Development. Luxembourg: Institute of Air Navigation Services, 2010.