

Perceptual and Performance Aspects of Flight Simulation: A Didactic Comparison of Projection-Based and Virtual Reality Display

Pavlina ENGELLOVA^{1*}, Vladimír SMRZ¹, Petr MICHENKA¹, Jiri ULVR¹, Ondřej MACH¹

¹Air Force Department, Faculty of Military Technology, University of Defence, Kounicova 65, 602 00, Brno.

Correspondence: *pavlina.engelova@unob.cz

Abstract

This pilot study evaluates projection-based and virtual reality visualization in flight simulation using objective performance metrics and NASA-TLX workload assessment. A within-subject design with eight pilot students was used to verify a segment-based evaluation framework. Results indicate that virtual reality supports more stable control with lower variability, while projection-based display enables higher accuracy relative to predefined targets. Workload remained low to moderate in both conditions. The findings suggest that visualization modality influences control strategy and support the suitability of the proposed methodology for further research and application in simulation-based pilot training.

KEY WORDS: *flight simulation, virtual reality, pilot performance, workload assessment, simulation training.*

Citation: Engelova, P.; Smrz, V.; Michenka, P.; Ulvr, J.; Mach, O. Perceptual and Performance Aspects of Flight Simulation: A Didactic Comparison of Projection-Based and Virtual Reality Display. 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.2026.15>

1. Introduction

The rapid advancement of immersive visualization technologies has significantly reshaped contemporary flight simulation and pilot training environments. Virtual reality (VR) systems based on head-mounted displays (HMDs) offer stereoscopic rendering, extended field of view, and real-time head tracking, features commonly associated with increased immersion and perceptual realism [1; 2; 3]. These characteristics have stimulated interest in VR as a potentially transformative instructional medium in aviation education. Nevertheless, projection-based visualization systems remain widely implemented in academic and certified training environments due to their perceptual stability, lower susceptibility to simulator sickness, and compatibility with established instructional procedures [4; 5].

A fundamental conceptual distinction must be made between immersion and presence. Immersion refers to objective technological characteristics of the system, including field of view, resolution, refresh rate, and latency. Presence, in contrast, represents the subjective psychological experience of “being there” within the virtual environment [1; 2]. Slater [3] emphasizes that immersive systems may create an illusion of presence without necessarily enhancing task-relevant cognitive processing. Consequently, increased immersion does not automatically translate into improved performance, particularly in operational domains requiring precise psychomotor control and sustained cognitive regulation.

The broader debate concerning simulation fidelity further complicates the interpretation of immersive technologies. Traditional assumptions equated higher physical fidelity with improved transfer of training. However, seminal work by Baldwin and Ford [6] demonstrated that transfer outcomes depend on multiple interacting variables, including trainee characteristics and work environment factors. Hays and Singer [7] argued that functional task alignment may be more critical than visual similarity, while Alexander et al. [8] highlighted the importance of psychological fidelity, defined as the degree to which the simulator evokes cognitive processes equivalent to those required in the operational setting. These perspectives suggest that technological immersion alone is insufficient as a predictor of training effectiveness.

From a cognitive standpoint, Sweller’s Cognitive Load Theory [9] provides a robust explanatory framework for interpreting the effects of immersive visualization. Learning efficiency depends on balancing intrinsic, extraneous, and germane cognitive load. Highly immersive environments may inadvertently increase extraneous load if perceptual stimulation exceeds task-relevant processing capacity.

Recent empirical findings suggest that immersive virtual environments may actively modulate fundamental perceptual mechanisms. Bogon et al. [10] demonstrated that immersive VR significantly alters time perception depending on spatial stimulus configuration, indicating that immersive systems can influence temporal and spatial processing at a basic cognitive level. Such perceptual modulation may indirectly affect attentional allocation and working memory demands.

This relationship between psychophysiological state and flight performance has been confirmed in aviation research, where changes in physiological and cognitive factors were shown to directly influence pilot performance and operational effectiveness [11].

In precision-based tasks such as aircraft control, where stable sensorimotor integration is required, excessive sensory input, visual latency, or perceptual distortion may reduce available cognitive resources and impair performance stability. The relationship between arousal and performance is further described by the Yerkes–Dodson law [12], which proposes an inverted-U relationship between activation and task effectiveness. Both under-stimulation and over-stimulation may degrade performance, making empirical comparison between visualization modalities essential.

Skill acquisition research also emphasizes structured feedback and measurable performance indicators as central to expertise development. Ericsson’s theory of deliberate practice [13] identifies systematic repetition, objective measurement, and performance feedback as key mechanisms of expert progression. In flight simulation, quantitative indicators such as root mean square error (RMSE) of control deviations align with these principles by providing objective metrics of precision and stability.

These principles imply that measurable and validated performance metrics are essential when evaluating training technologies. Within aviation specifically, Allerton [5] stresses the necessity of validation and verification of flight simulation devices to ensure training relevance and methodological credibility. The introduction of emerging technologies such as VR into academic flight training therefore requires systematic evaluation using validated performance metrics and established human factors tools.

Workload assessment represents an essential complement to objective performance evaluation. The NASA Task Load Index (NASA-TLX) is one of the most widely validated multidimensional workload instruments in aviation research [14; 15]. However, comprehensive workload assessment in aviation increasingly emphasizes the integration of subjective and physiological indicators, as physiological parameters provide additional insight into the operator’s psychophysiological state and its relationship to performance [16]. Performance cannot be interpreted independently of perceived task demand, as workload mediates attentional allocation, error rates, and psychomotor stability. Kennedy et al. [4] demonstrated that simulator-induced discomfort can significantly influence operational performance, reinforcing the need for integrated performance–workload assessment frameworks.

Meta-analytic evidence confirms that virtual reality-based training programs can produce meaningful learning effects. However, their effectiveness depends strongly on instructional design, task specificity, and methodological rigor. [17] Immersive technology alone does not guarantee superior transfer.

In the context of academic military aviation education, simulator-based performance monitoring has demonstrated predictive value. Ulvr et al. [18] reported a moderate correlation between simulator exercise performance and real flight training outcomes among military pilot students. Their findings support the structured use of simulator-derived metrics for competency development tracking and curriculum optimization.

Similarly, previous research has demonstrated that simulator-based training can significantly influence piloting precision and control quality, further supporting its role as a valid tool for performance-oriented training and assessment [19].

Taken together, existing literature suggests that immersive visualization may influence flight performance indirectly through perceptual-cognitive modulation and workload mediation rather than through visual realism alone.

Despite growing interest in immersive systems, there remains limited empirical research directly comparing projection-based and VR-based visualization using objective flight control accuracy metrics combined with validated workload assessment.

The present study addresses this gap by implementing a structured within-subject experimental comparison of projection-based and VR visualization modalities in an academic flight simulation environment. The study integrates objective control accuracy metrics, multidimensional workload assessment, and theoretical grounding in transfer of training and cognitive load frameworks. Rather than aiming to demonstrate technological superiority, the primary objective is to validate a performance-oriented evaluation model suitable for academic aviation research and competency-based training environments.

2. Methods

The empirical investigation was designed as a controlled within-subject pilot experiment conducted in an academic flight simulation laboratory. The within-subject design was selected to reduce inter-individual variability in baseline piloting skills, perceptual sensitivity, and cognitive processing capacity, thereby increasing statistical power in a limited exploratory sample [20].

Each participant completed identical standardized flight tasks under two visualization conditions:

1. Projection-based multi-screen display system.
2. Virtual reality head-mounted display (HMD).

Each participant completed the same predefined flight task in both visualization conditions, and the resulting datasets were paired at the subject level. The analytical comparison therefore reflected within-subject differences between the projection-based and virtual reality conditions rather than between-group variation.

The visualization modality served as the independent variable. Primary dependent variables included objective flight control accuracy and subjective workload. This design directly reflects theoretical assumptions that immersion may influence performance indirectly via perceptual-cognitive and workload mechanisms rather than solely through visual realism [9; 10; 18].

To control for order and learning effects, the sequence of visualization conditions was counterbalanced. Participants were randomly assigned to one of two sequences:

1. Sequence 1: Projection → VR.
2. Sequence 2: VR → Projection.

A short familiarization phase preceded each condition to minimize novelty effects and perceptual adaptation bias.

2.1. Experimental Environment and Apparatus

The experiment was conducted using a fixed-base cockpit simulator replicating a light aircraft configuration (Fig. 1). The simulator included:

1. Fully functional flight control hardware.
2. Realistic avionics instrumentation.
3. Identical cockpit ergonomics in both conditions.



Fig. 1. Experimental flight simulation setup.

In the projection-based condition, the outside visual scene was rendered on a wide-field projection screen providing peripheral contextual cues.

In the VR condition, the same simulation environment was presented using a stereoscopic HMD with head tracking enabled. System latency and rendering parameters complied with manufacturer specifications to minimize simulator sickness risk [4].

Aircraft flight model, atmospheric conditions, wind parameters, and environmental settings were identical across both visualization modalities. No turbulence, failures, or stochastic disturbances were introduced in order to isolate visualization effects. Flight data were recorded directly from the simulator environment as time-series outputs and subsequently processed offline. The exported dataset included positional, kinematic, and attitude-related variables, particularly longitude, latitude, altitude, heading, vertical speed, roll, pitch, yaw, and flight-path angle. The same simulator configuration, route structure, and environmental settings were used in both visualization conditions to preserve comparability of the recorded performance data.

The emphasis on controlled comparability aligns with established principles of simulation fidelity evaluation [6; 7].

2.2. Participants

The experimental sample consisted of eight participants, all of whom were students enrolled in the Military Pilot specialization at the University of Defence. The selection of this group was intentional, as all participants represented a relatively homogeneous population with comparable academic background, training context, and expected familiarity with fundamental aviation procedures. Such homogeneity is methodologically advantageous in a pilot within-subject design, as it reduces the influence of uncontrolled inter-individual variability unrelated to the visualization modality under investigation.

All participants had prior exposure to simulator-based training and possessed theoretical knowledge of basic aircraft control, including attitude stabilization, heading control, climb and descent management, and coordinated maneuvering. Their

previous instructional background ensured that the recorded performance reflected adaptation to the visualization environment rather than a first-contact response to elementary flight tasks.

Eligibility for participation was defined by several practical and safety-related criteria. Participants were required to have normal or corrected-to-normal vision, no diagnosed vestibular disorder that could increase susceptibility to visually induced discomfort, and basic familiarity with flight simulator controls and cockpit task execution. These criteria were introduced to minimize confounding factors associated with visual limitation, motion sensitivity, or insufficient procedural preparedness.

The study was conducted in accordance with institutional ethical standards applicable to educational and experimental activities in the academic training environment. Participation was voluntary, and the investigation was designed as a pilot study with the primary aim of methodological verification of the proposed evaluation framework rather than population-level inference. Consequently, the participant group should be understood as an exploratory academic sample suitable for initial comparison of projection-based and virtual reality display conditions in a controlled flight simulation setting.

2.3. Experimental Task

The experimental task consisted of a structured flight profile with an approximate duration of 40 minutes, designed to reflect a sequence of maneuver-based training elements organized along a predefined spatial trajectory. Unlike conventional time-driven task designs, the flight scenario was defined by a planned route consisting of seven waypoints, resulting in six analytically distinct flight segments. Segment boundaries were determined automatically from waypoint-passage detection based on the recorded spatial trajectory. For each subject and condition, the nearest passage to each predefined waypoint was identified in chronological order, and the intervals between successive waypoint passages were treated as individual analytical segments.

Each segment corresponded to a specific maneuver type and performance requirement, as defined in the experimental scenario and subsequently used in the data processing pipeline for segment-based evaluation. The segmentation was not imposed temporally but was derived from actual waypoint passage detection, ensuring alignment between intended and executed flight structure.

The task included the following maneuver categories:

- 1. Segment 1 – Low-altitude maneuvering:**
Maintaining approximately 1000 ft AGL while performing horizontal 180° turns with a 45° bank to both left and right, followed by controlled pitch maneuvers consisting of a 10° climb and a 10° descent, each sustained for 30 seconds.
- 2. Segment 2 – Combined vertical and turning maneuvers:**
Execution of a climbing left turn (360°, 45° bank, +10° pitch) followed by a descending right turn (360°, 45° bank, -10° pitch), requiring simultaneous coordination of pitch and roll.
- 3. Segment 3 – Low-altitude control with aerobatic elements:**
Maintaining approximately 1000 ft AGL while performing roll maneuvers, increasing demands on attitude control and stability.
- 4. Segment 4 – Terrain-following climb:**
Flight through a valley profile combined with a continuous climb to a target altitude of approximately 12,000 ft AMSL.
- 5. Segment 5 – Dynamic maneuvering task:**
Aircraft pursuit scenario requiring continuous trajectory adaptation and relative motion control.
- 6. Segment 6 – Descent and complex maneuvering:**
Descent to approximately 1000 ft AGL followed by slalom flight between pylons and repeated horizontal 180° turns with a 45° bank to both sides.

This structure resulted in a heterogeneous task profile, combining steady-state control, transient maneuvering, and continuous control adjustments. Such design enables differentiation between precision control, dynamic stability, and pilot adaptation to changing task demands.

The spatial trajectory recorded during the experiment confirms that the task induced recurrent maneuvering patterns, including localized circular trajectories near specific waypoints. These patterns correspond to instructed maneuver types such as turns and spirals, rather than deviations or navigation errors. Importantly, differences between visualization conditions (projection vs VR) are observable not only in trajectory dispersion but also in the smoothness and spatial consistency of maneuver execution, particularly in high-workload segments.

Target values were not defined as a continuous time-dependent reference function but rather as segment-specific performance objectives, including:

1. Target altitude (e.g., climb segment).
2. Target pitch magnitude.
3. Target bank angle.
4. Target flight path angle.

This approach reflects realistic flight training practice, where pilots are required to achieve and maintain discrete flight states rather than continuously track an externally defined trajectory.

From a cognitive perspective, the task was intentionally designed to introduce varying levels of control demand across segments, thereby modulating intrinsic task complexity. This aligns with Cognitive Load Theory, which emphasizes that learning-relevant performance differences emerge when task difficulty is appropriately matched to operator capacity [9].

The inclusion of maneuver-intensive segments (e.g., spirals) alongside stabilized phases enables observation of performance under both moderate and elevated workload conditions.

Finally, the use of a route-based segmentation framework combined with maneuver-specific task definitions ensures that the experimental design supports direct linkage between:

1. Control behavior.
2. Task type.
3. Visualization modality.

This structure is particularly suitable for subsequent segment-level performance analysis and target deviation evaluation, as implemented in the data processing pipeline.

2.4. Performance Metrics and Operationalization

In contrast to classical continuous tracking approaches based on a time-dependent reference signal $x_{ref}(t)$, the present study adopts a segment-based evaluation framework without a continuous reference trajectory. For each subject and condition, the analysis was performed both globally over the full flight and locally within route-defined segments. This methodological choice reflects the structure of the experimental task, which is defined by discrete maneuver segments and target flight states rather than continuous tracking requirements.

Accordingly, performance was not evaluated using traditional time-series error metrics such as Root Mean Square Error (RMSE) or Mean Absolute Error (MAE), but instead through a combination of:

1. Descriptive statistical indicators.
2. Segment-specific target deviation metrics.

2.5. Global Statistical Descriptors

For each flight parameter x , evaluated over a set of samples $\{x_i\}_{i=1}^n$, the following descriptors were computed:
Mean value:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

Standard deviation:

$$\sigma_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

Range:

$$R_x = \max(x_i) - \min(x_i) \quad (3)$$

These metrics provide a compact characterization of control behavior in terms of:

1. Central tendency (trim behavior).
2. Variability (stability).
3. Amplitude of deviations (control envelope).

Unlike RMSE-based approaches, this formulation does not assume the existence of a predefined optimal trajectory and is therefore better aligned with instructional flight tasks based on state maintenance rather than trajectory tracking.

2.6. Segment-Level Evaluation

Let the flight be partitioned into S segments defined by waypoint passage detection. For each segment s , a subset of samples $\{x_i^{(s)}\}$ is extracted, and the same statistical descriptors are computed:

$$\bar{x}^{(s)}, \sigma_x^{(s)}, R_x^{(s)} \quad (4)$$

This segmentation enables task-specific performance analysis, allowing differentiation between:

1. Stabilized flight phases.
2. Maneuver-intensive segments (turns, spirals).
3. Transition regimes (climb/descent).

Such structure is essential for identifying performance modulation as a function of task complexity, rather than averaging behavior over the entire flight.

2.7. Target-Based Deviation Metrics

For segments with explicitly defined performance objectives (e.g., target altitude, pitch, roll, or flight path angle), deviation from target value x_{target} was quantified using absolute deviation metrics.

Mean absolute deviation from target:

$$E_{\text{mean}}^{(s)} = \frac{1}{n_s} \sum_{i=1}^{n_s} |x_i^{(s)} - x_{target}| \quad (5)$$

Final state deviation:

$$E_{\text{final}}^{(s)} = x_{n_s}^{(s)} - x_{target} \quad (6)$$

For variables defined in terms of magnitude (e.g., roll angle or pitch angle), deviation was evaluated as:

$$E_{\text{mag}}^{(s)} = \frac{1}{n_s} \sum_{i=1}^{n_s} | |x_i^{(s)}| - x_{target} | \quad (7)$$

Importantly, when comparing performance quality, signed errors were evaluated using their absolute magnitude, ensuring that deviation direction does not bias the interpretation of control accuracy.

2.8. Comparison Between Visualization Conditions

Performance comparison between projection-based (VIS) and virtual reality (VRG) conditions was conducted using paired metric comparison.

For each segment s , variable x , and metric m , the difference was defined as:

$$\Delta_{s,x,m} = x_{VRG}^{(s,m)} - x_{VIS}^{(s,m)} \quad (8)$$

A lower value was interpreted as better performance for:

1. Variability metrics (standard deviation).
2. Range.
3. Target deviation metrics.

This formulation enables direct comparison without requiring normalization or transformation of the original data.

2.9. Winner-Based Aggregation

To enhance interpretability at both segment and group level, a winner-based aggregation was introduced as an auxiliary descriptive indicator. For each segment, variable, and metric, the condition with the lower value (i.e., lower variability or lower deviation) was labeled as the “winner.”

The number of favorable outcomes for each condition was then counted across metrics, resulting in a simplified summary of directional differences:

$$W_{VIS}(s), W_{VRG}(s).$$

This approach does not quantify the magnitude of differences but provides an intuitive representation of the consistency and direction of performance differences across multiple variables and task segments.

All metrics were treated with equal weight in this aggregation, and no normalization across variables was applied.

The winner-based summary should therefore be interpreted as a complementary descriptive tool and not as a substitute for continuous performance evaluation.

2.10. Data Preprocessing

Prior to analysis, all time series were sorted according to relative time. Only valid samples were retained for each evaluated variable, and missing values were excluded on a variable-by-variable basis. In datasets containing multiple simulator identifiers, only records corresponding to the target aircraft were retained. No smoothing or filtering was applied, as preserving natural control variability was considered essential for the interpretation of pilot behavior.

2.10.1. Theoretical Justification.

The adopted performance framework is grounded in three complementary theoretical perspectives. First, from a control theory standpoint, system performance is not solely defined by tracking error relative to an external reference but also by stability, variability, and control effort, which are effectively captured by dispersion-based metrics [21].

Second, the approach aligns with the deliberate practice framework, where performance improvement is driven by structured tasks and measurable outcomes rather than continuous trajectory tracking [13].

Third, within human factors research, variability and stability of control inputs are closely associated with cognitive workload, attentional allocation, and perceptual processing constraints [14; 15]. Therefore, the selected metrics provide indirect insight into underlying cognitive mechanisms influencing pilot performance.

By combining statistical descriptors, target-based deviations, and comparative evaluation across task segments, the framework enables a multidimensional characterization of flight performance. This is particularly relevant when evaluating immersive visualization technologies, where performance differences may manifest not only as changes in average accuracy but also in stability, consistency, and control dynamics.

Workload Assessment. Subjective workload was assessed after completion of the experimental task using the NASA Task Load Index (NASA-TLX) in its raw, unweighted form [14; 15]. Participants rated six workload dimensions, including mental demand, physical demand, temporal demand, perceived performance, effort, and frustration.

Each dimension was rated on a 0–100 scale. The Performance dimension was reverse-coded ($100 - \text{Performance}$) so that higher values consistently indicated higher workload across all dimensions. The overall workload score (TLX Raw Mean) was computed as the arithmetic mean of the six dimensions using the reversed Performance score. The NASA-TLX was administered once after completion of the entire experimental session and therefore reflects overall perceived workload rather than condition-specific workload. The NASA-TLX has demonstrated strong reliability and construct validity in aviation and human performance research [14; 15].

2.10.2. Statistical Analysis

Given the exploratory nature of the study and the limited sample size ($n = 8$), the analysis was based on descriptive and comparative evaluation rather than inferential statistical testing.

Performance differences between visualization conditions (VIS vs VRG) were evaluated using:

1. Direct comparison of computed metrics.
1. Segment-level aggregation across participants.
2. Winner-based dominance analysis.

For each metric, differences between conditions were expressed as:

$$\Delta = x_{VRG} - x_{VIS} \quad (10)$$

and interpreted in terms of direction and magnitude rather than statistical significance.

Group-level results were obtained by averaging metrics across participants and by evaluating the consistency of observed effects within segments. The winner-based approach further enabled intuitive classification of condition dominance at both segment and group level. No hypothesis testing (e.g., t-tests) was performed due to the small sample size and the pilot character of the study. Instead, emphasis was placed on consistency of effects, task-specific patterns, and agreement between objective performance and subjective workload. This approach is consistent with methodological recommendations for exploratory studies in applied human performance research.

3. Results

A total of eight subjects completed the experimental task under both display conditions (VIS and VRG). The results are presented at four levels: trajectory overview, global and segment-based objective performance metrics, target-deviation measures, and subjective workload assessed using NASA-TLX.

3.1. Trajectory overview

The aggregate trajectory plots (see Fig. 2 and 3) show that all subjects in both conditions followed the planned route structure and waypoint sequence. The largest spatial dispersion occurred in maneuvering regions, particularly during turning segments and repeated heading changes, while straight flight legs remained relatively compact.

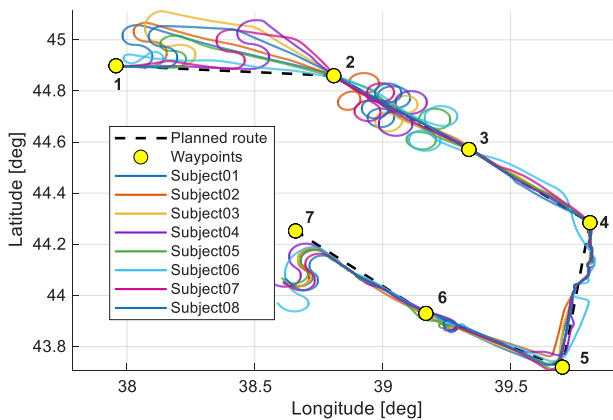


Fig.2. Aggregate trajectories of all subjects in the VIS condition.

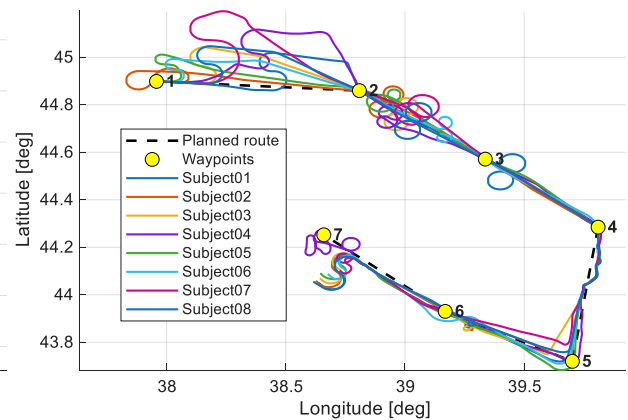


Fig. 3. Aggregate trajectories of all subjects in the VRG condition.

No qualitative differences in route topology between VIS and VRG were observed. Both conditions produced comparable spatial execution patterns. The trajectory visualization therefore serves primarily as a spatial reference for subsequent quantitative analysis.

3.2. Global performance metrics

Global descriptive statistics across the full flight are summarized in Table 1.

Table 1.

Global descriptive metrics across the full flight (VIS vs. VRG).

Variable	VIS (Mean ± Std)	VRG (Mean ± Std)
Altitude	1416 ± 1454	1286 ± 1343
FlightPath	0.074 ± 6.75	0.079 ± 5.88
Heading	182.4 ± 86.6	181.7 ± 86.3
Pitch	2.11 ± 6.81	2.20 ± 5.96
Roll	0.23 ± 31.9	1.45 ± 30.1
VS	~0 ± 22.9	~0 ± 20.2
Yaw	185.7 ± 86.7	185.0 ± 86.4

Across all evaluated variables, mean values were comparable between VIS and VRG, indicating a similar overall level of task execution. In contrast, variability-related metrics (range and standard deviation) showed more consistent differences between conditions. Specifically, VRG demonstrated lower variability in most kinematic and attitude parameters, including altitude, flight-path angle, pitch, roll, and vertical speed. This pattern suggests that differences between conditions were primarily reflected in the dispersion of the data rather than in absolute mean values.

Detailed numerical results for the global descriptive statistics reported in this section are provided in Table 1.

3.3. Segment-based performance

Representative segment-level variability results for altitude are summarized in Table 2 and provide a more detailed view of task execution across individual maneuver phases. The comparison between VIS and VRG indicates that performance differences were not uniform across segments. Instead, a consistent pattern of segment-dependent variability emerged. Segment 1 showed a balanced outcome between conditions, with comparable variability levels. In contrast, Segments 2, 4, 5, and 6 exhibited systematically lower variability in VRG across the evaluated parameters. Segment 3 demonstrated a mixed behavior, with higher dispersion in altitude in VRG, while several attitude-related variables showed reduced variability.

Table 2.

Segment-based variability (Altitude).

Segment	VIS (Mean ± Std)	VRG (Mean ± Std)
S1	445 ± 230	430 ± 245
S2	1732 ± 1616	1251 ± 1160
S3	600 ± 103	728 ± 269
S4	2044 ± ~800	1939 ± ~650
S5	3707 ± 429	3476 ± 484
S6	691 ± 692	667 ± 575

Overall, the results indicate that performance differences between conditions were strongly segment-dependent, with a prevailing trend of reduced variability in VRG across most segments.

3.4. Target-deviation metrics

Target-based evaluation results are summarized in Table 3.

Table 3.

Target-deviation metrics for selected segments (mean values).

Segment	Metric	Target	VIS	VRG
S1	Flight path error [deg]	0	8.25	8.17
S1	Pitch error [deg]	0	7.27	7.32
S1	Roll error [deg]	0	31.20	31.31
S2	Flight path error [deg]	±10	4.47	5.72
S2	Pitch error [deg]	±10	4.83	5.50
S2	Roll error [deg]	45	21.17	25.35
S4	Mean altitude error [ft]	12000	9956	10061
S4	Final altitude error [ft]	12000	-8356	-8329
S6	Roll error [deg]	45	30.76	31.84

Mean values are reported. Variability measures were not available for target-based metrics.

In contrast to the descriptive segment metrics, the target-deviation analysis revealed a different pattern. Segment 1 showed negligible differences between conditions. Segment 2 demonstrated consistently lower deviation values in VIS, particularly for attitude-related variables. Segments with endpoint constraints, such as altitude targets, exhibited minimal differences between conditions, with a high proportion of ties at the subject level.

3.5. Subject-level summary

The winner-count analysis provides a compact within-subject comparison of performance between VIS and VRG across all evaluated metrics. It should be noted that this aggregation reflects only the direction of differences and does not capture their magnitude.

For segment-based descriptive metrics, the winner-based summary indicated a higher frequency of favorable outcomes for VRG. As shown in Table 4, the majority of subjects achieved a higher number of lower-metric outcomes in VRG, indicating a higher frequency of lower-variability outcomes across segments.

Table 4.

Subject-level winner summary for segment-based descriptive metrics.

Subject	VIS wins	VRG wins	Overall winner
S01	18	66	VRG
S02	24	60	VRG
S03	30	54	VRG
S04	52	32	VIS
S05	18	62	VRG
S06	8	72	VRG
S07	28	56	VRG
S08	26	58	VRG

Table 5.

Subject-level winner summary for target-based metrics.

Subject	VIS wins	VRG wins	Overall winner
S01	1	0	VIS
S02	1	0	VIS
S03	1	0	VIS
S04	1	0	VIS
S05	1	0	VIS
S06	1	0	VIS
S07	0	1	VRG
S08	0	1	VRG

In contrast, the target-based evaluation showed an opposite trend. As summarized in Table 5, most subjects demonstrated lower deviation values in VIS, indicating higher accuracy relative to predefined targets.

This divergence between segment-based variability and target-based accuracy highlights that the two display conditions differed not only in overall performance, but also in the nature of control behavior.

3.6. NASA-TLX workload

NASA-TLX results are summarized in Table 6 and illustrated in Fig. 4 and Fig. 5.

Table 6.

Descriptive statistics of NASA-TLX scores across subjects.

Metric	Mean	SD	Min	Max	N
Mental Demand	31,88	26,45	5	75	8
Physical Demand	14,38	6,78	5	25	8
Temporal Demand	25,63	21,45	5	65	8
Performance	70,63	18,98	35	90	8
Performance_rev	29,38	18,98	10	65	8
Effort	30,63	18,60	15	65	8
Frustration	25,63	18,21	10	60	8
TLX Raw Mean	26,25	15,70	10,83	57,5	8

At the group level, the workload profile was characterized by the highest mean values in Performance, followed by Mental Demand, Effort, and Performance_rev. Physical Demand showed the lowest mean score, while Mental Demand exhibited the highest inter-individual variability.

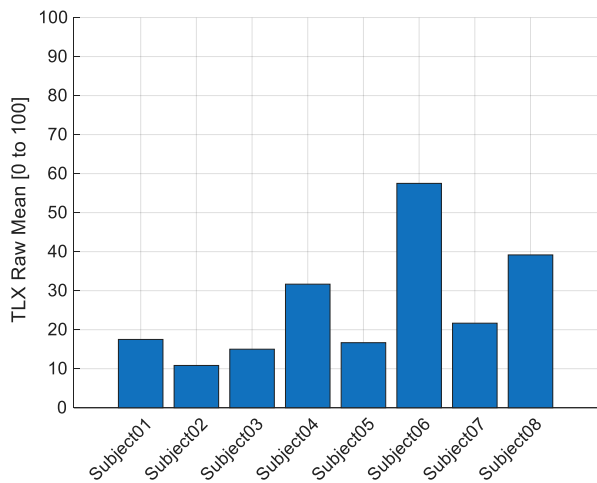


Fig. 4. NASA-TLX composite workload by subject.

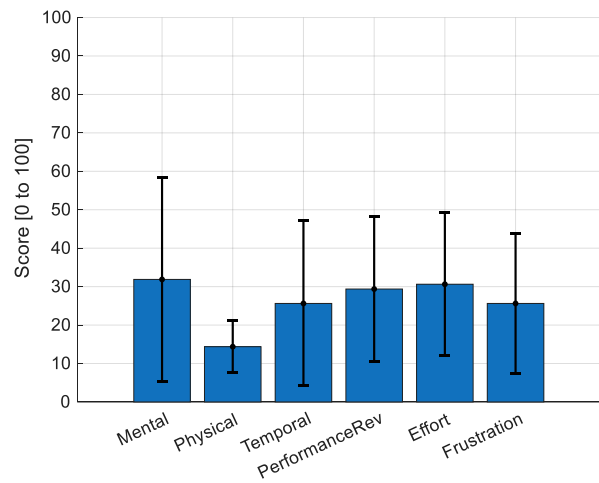


Fig. 5. NASA-TLX dimensional profile across subjects.

The composite NASA-TLX score, expressed as the raw mean across dimensions, indicated an overall low-to-moderate subjective workload across subjects. At the same time, substantial inter-individual variability was observed, with some subjects reporting very low workload and others reaching values near the upper end of the observed range. Among the individual dimensions, Mental Demand exhibited the highest variability, whereas Physical Demand remained the lowest-rated dimension across subjects. Detailed numerical values for all dimensions and subject-level scores are provided in Table 6.

4. Discussion

The results reveal a consistent divergence between variability-based and target-based performance metrics across the two display conditions. From the segment-based perspective, VRG was associated with lower variability across most segments and subjects, indicating more stable control behavior. This pattern was also reflected in the winner-based aggregation, which indicated a higher frequency of lower-variability outcomes in VRG. However, this representation reflects only the direction of differences and does not capture their magnitude. In contrast, the target-deviation analysis demonstrated higher accuracy in VIS. These findings suggest that the two visualization modalities are associated with different control characteristics, with VRG linked to reduced variability and VIS to higher target accuracy. These differences may be related to variations in perceptual and control strategies; however, the underlying perceptual-cognitive mechanisms were not directly measured in this study and therefore remain hypothetical. In this context, the observed patterns are consistent with theoretical perspectives emphasizing functional task alignment over visual realism [7; 8].

The NASA-TLX results indicate an overall low-to-moderate subjective workload during the experimental task. This suggests that the observed performance differences were not associated with pronounced subjective overload. At the same time, substantial inter-individual variability was observed, particularly in mental demand. It should be noted that workload was assessed globally after completion of the task and was not measured separately for individual visualization conditions.

From a human factors perspective, the observed variability patterns may be interpreted in relation to the Yerkes–Dodson law [12], where different levels of stimulation may favor either stability or precision; however, this interpretation remains indirect. From a training perspective, VRG may be suitable for developing control stability and coordination, whereas VIS may be more effective for precision tasks requiring adherence to defined flight parameters. This supports the view that the effectiveness of simulation technologies depends on task specificity rather than immersion alone [17].

The study is limited by a small sample size and the absence of variability measures in target-based metrics. In addition, the winner-based aggregation reflects only the direction of differences and should be interpreted together with continuous performance metrics and effect-size estimates. Future research should include larger samples, additional performance indicators, and direct measurement of perceptual and cognitive processes underlying control behavior.

5. Conclusions

This study compared projection-based and virtual reality visualization in a controlled flight simulation task using objective performance metrics and NASA-TLX workload assessment. The results indicate a consistent divergence between performance dimensions. VRG was associated with lower variability across most segments, as reflected in both continuous metrics and the winner-based aggregation, whereas VIS demonstrated higher accuracy relative to predefined task targets. These findings suggest that the two visualization modalities are associated with different control characteristics rather than one being universally superior. The overall subjective workload was low to moderate, indicating that the observed performance differences were not associated with pronounced perceived workload and cannot be explained solely by self-reported workload.

From a training perspective, VRG may be suitable for developing control stability and coordination, whereas VIS may be more effective for precision-based tasks requiring strict adherence to flight parameters. Overall, the study supports

the need for task-specific selection of visualization modalities and demonstrates the applicability of structured performance-based evaluation frameworks in simulation-based pilot training.

The present study addresses this gap by implementing a structured within-subject experimental comparison of projection-based and VR visualization modalities in an academic flight simulation environment, integrating objective control accuracy metrics, multidimensional workload assessment, and theoretical grounding in transfer of training and cognitive load frameworks. Rather than aiming to demonstrate technological superiority, its primary objective is to validate a performance-oriented evaluation model suitable for academic aviation research and competency-based training environments, which is subsequently confirmed by the results of the structured experimental comparison.

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

References

1. **Witmer B.G., Singer M.J.** Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 1998, 7, p. 225-240.
2. **Slater M.** Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 1999, 8, p. 560-565.
3. **Slater M.** Immersion and the illusion of presence in virtual reality. *British Journal of Psychology*, 2018, 109, p. 431-433.
4. **Kennedy R.S., Lane N.E., Berbaum K.S., Lilienthal M.G.** Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, 1993, 3, p. 203-220.
5. **Allerton D.** *Principles of Flight Simulation*. 2009, Chichester: John Wiley.
6. **Baldwin T.T., Ford J.K.** Transfer of training: A review and directions for future research. *Personnel Psychology*, 1988, 41, p. 63-105.
7. **Hays R.T., Singer M.J.** *Simulation Fidelity in Training System Design: Bridging the Gap Between Reality and Training*. New York: Springer, 2012.
8. **Alexander A., Bruny  T., Sidman J., Weil S.** From Gaming to Training: A Review of Studies on Fidelity, Immersion, Presence, and Buy-in and Their Effects on Transfer in PC-Based Simulations and Games. *Simulation & Gaming*, 2005.
9. **Sweller J.** Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 1988, 12, p. 257-285.
10. **Bogon J., Jagorska C., Steinecker I., Riemer M.** Age-related changes in time perception: Effects of immersive virtual reality and spatial location of stimuli. *Acta Psychologica*, 2024, 249, 104460.
11. **Zgodavov Z., Rozenberg R., Szabo S., Sabo J.** The Impact of Changes of Psychophysiological Factors on the Flight Crew Performance. *International Review of Aerospace Engineering*, 2019, 12 (3).
12. **Yerkes R.M., Dodson J.D.** The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 1908, 18, p. 459-482.
13. **Ericsson K.A., Krampe R.T., Tesch-R mer C.** The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 1993, 100, p. 363-406.
14. **Hart S.G., Staveland L.E.** Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology*, vol. 52. Elsevier, 1988, p. 139-183.
15. **Rubio S., Daz E., Martn J., Puente J.M.** Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Applied Psychology*, 2004, 53, p. 61-86.
16. **Hanakova L., Socha V., Socha L., Szabo S., Kozuba J., Lalis A.** Determining importance of physiological parameters and methods of their evaluation for classification of pilots psychophysiological condition. In *Proceedings of 2017 International Conference on Military Technologies (ICMT)*. IEEE, 2017, p. 500-506.
17. **Howard M.C., Gutworth M.B., Jacobs R.R.** A meta-analysis of virtual reality training programs. *Computers in Human Behavior*, 2021, 121, 106808.
18. **Ulvr J., Smrz V., Mach O.** Monitoring the Development of Military Pilot Competencies During Studies at the University of Defence. In *Proceedings of New Trends in Aviation Development (NTAD)*, 2024, p. 205-211.
19. **Socha V., Socha L., Szabo S., Hana K., Gazda J., Kimlickova M.** Training of pilots using flight simulator and its impact on piloting precision. *Transport Means*, 2016, p. 374-379.
20. **Cohen J.** *Statistical Power Analysis for the Behavioral Sciences*. New York: Routledge, 1969.
21. **Holmes A.** *Modern Control Engineering*. Murphy & Moore Publishing, 2021.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of CNDCGS 2026 and/or the editor(s). CNDCGS 2026 and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.