

DISTRIBUTIONS OF AIR TRAFFIC CONTROL STUDENTS' ATTITUDES TOWARDS WORKLOAD

Serhii BORSUK ^{1*}, Oleksii REVA ²

¹*Institute of Laser and Optoelectronics Intelligent Manufacturing, Wenzhou University, Wenzhou, China*

²*Ukrainian Institute of Scientific and Technical Expertise and Information, Kyiv, Ukraine*

Received 10 November 2020; accepted 8 May 2021

Abstract. Mental workload is a well-known concept with a long development history. It can be used to examine students' attitudes at the end of the educational process and compare them in groups or separately. However, building a continuous workload profile across the range of task complexity increase is still an urgent issue. All four groups of methods used to define mental workload have such flaws for the workload profile construction process as significant time requirements, single value processing and post-processing of the received results. Only one of them can be used without modifications to construct the operator's attitude chart (profile) regarding the workload range and it doesn't operate with more reliable absolute values. To resolve this problem, a special workload assessment grid was developed, considering the advantages of a subjective group of methods and seven core characteristics. The reasoning for grid axes choice, threshold values, and question formulation were provided. Statistics were calculated for the full sample, different grades, and educational institutions. Comparison of the received responses with referential values, cross-comparison between institutions and different grades were performed. The results contribute to such important aspects of workload, as redlines, workload profiling, and operator's comparison.

Keywords: human factors, air traffic control, education, workload, self-assessment, Yerkes-Dodson law.

Introduction

Mental workload active research and examination last since the 1960s (Yan et al., 2019). The Mental workload (MWL) concept was developed as a response to challenges posed by the industry in terms of rapid technology performance growth and workplace saturation with additional information and data. Along with a concept of situation awareness, MWL is mainly intended to explain the operator's task performance and his orientation in the complex workplace environment (Chiappe et al., 2016). In general, MWL is closely connected to performance (Nachreiner, 1995; Stanton, 2005), and it might influence human errors, thus increasing safety. Taking into account a share of errors caused by operators (70%, according to Leveson, 2004), MWL analysis and manipulation are tasks of high importance.

There are many different definitions of MWL. Gao et al. (2013) define it as an amount of mental effort required to perform a particular task by an operator. Stanton (2005) claims that MWL has no single correct definition and consists of stress and strain components in general. Tsang and Vidulich (2012) mention debates since the

1970s and refer to Wickens, who describes MWL as a construct with quantity prevailing over quality (Wickens, 2002). Rusnock and Borghetti (2018) define MWL as an operator's effort value during task performance. Notwithstanding the particular terms used by the authors, they all agree on MWL high complexity, multidimensionality, quantitative nature, and the central role of the operator. Many distinct factors constitute the integral MWL concept. They are stress, fatigue, motivation level, operator's skill, attention, boredom, and others. In aviation, MWL is commonly measured for the cabin crew and air traffic controllers (Efthymiou et al., 2019).

According to Rehmann (1995), all MWL measurement methods can be separated into four groups: performance, subjective, physiological, and analytical (although some methods possess features of several groups at the same time (Socha et al., 2020)). The first three groups include MWL assessment methods that directly address operators and are used with their active participation. The latter group mostly includes predictive methods applied beforehand. The performance group is dedicated to primary and secondary task observation (Stanton, 2005). These

*Corresponding author. E-mail: greyone.ff@gmail.com

procedures commonly measure the number of active tasks completed by the operator with given different time periods, intensity, proportion, and other group details. The subjective group includes processing of multiple questionnaires of a different kind or interviews (Wang et al., 2020). They are used to extract the internal MWL esteem produced by the operator. The physiological group focuses on the tangible values provided by the human body (Miller, 2001). These parameters include pupil diameter, heartbeat rate, electrooculogram values, and others.

Among all existing methods, some particular ones like ISA (Jordan, 1992) were developed specifically for ATC estimation. ISA belongs to the subjective MWL assessment group; it is somewhat simple to use, and its results were found to correlate significantly with the post-task MWL ratings. Another example is the MACE method (Goillau & Kelly, 1997), which uses relative indexes and also belongs to the subjective group. The authors also address Hopkin (1995), who states that common practice in ATM research is to use ISA and SWAT during the work or performing NASA-TLX as the work ends.

The importance of MWL assessment methods application in the ATC area is supported by multiple pieces of research dedicated to ATC risk management. In general, the operator's MWL in complex systems (Braarud, 2020; Yan et al., 2019) is an urgent area of research, and it is especially important for processes that concern vital risks. ATC belongs to exactly this type of process, bringing together risks, MWL, and ergonomic methods (Zeier, 1994).

1. Problem statement

The task is stated as follows: to analyze the attitude towards different MWL levels for several academic groups of ATC students mastering the same curriculum in different higher educational institutions and different grades via their MWL self-estimation. MWL level was defined as a number of aircraft under the operator's control simultaneously (Goillau & Kelly, 1997) without defining the density type (Rahman et al., 2016). MWL esteems were decided to be taken as a continuous set on the range of aircraft under control. The following steps were included in the process: method selection, attitude data collection, statistical parameters calculation and preliminary processing, general peculiarities determination, conclusions on different grades, and conclusions on different education institutions.

2. The two-side continuous scale method

2.1. Existing methods

For proper data collection, the suitable method of MWL estimation is required. According to the task statement, there was no necessity in a physiological group. Those methods are useful for pure body response to stress measurement. Meanwhile, the task required rather behavioral answers. Analytical methods were also rejected since no

predictions required during the task performance. The further choice between task-based performance and subjective methods in favor of the latter ones was due to the survey conditions.

The program used for participants' training had several scenarios, but the number of aircraft under simultaneous control was limited. It was rather difficult to create a new scenario, upload it into the system in every workplace, and collect the on-line attitude data while MWL was changing. Moreover, scenarios often require ATC student interaction with the instructor who plays the role of a pilot in contact. Verbal interaction inclusion in the process would require participants' number decrease or enormous time to complete. These and other less important considerations caused the reject of task-based methods. Implementation ease, administration simplicity, and low cost of subjective methods were also considered as a significant virtue. Finally, it was decided to rely on a simple principle: if you want to know the attitude (opinion), you should ask.

While searching across the methods of subjective MWL estimation, we found out that a surprisingly small number of them are applicable. It was planned to perform our MWL attitude analysis over the range of aircraft numbers. However, many methods were intended to process a single task overall MWL level (thus referring to the maximal value and ignoring the intermediate ones). This subgroup includes a framework proposed by Bommer (2016), NASA-TLX, SWORD (Vidulich et al., 1991), and SWAT. These methods also often require much time to get data processed correctly (Valdehita et al., 2004). Some methods can construct a continuous set of MWL rates. This subgroup includes ISA, the "Bedford" MWL scale (Roscoe & Ellis, 1990), PUMA (Kirwan et al., 1997), etc. There are many other methods listed and explained in various proceedings (Stanton, 2005; Rehmann, 1995; Tsang & Velazquez, 1996; Miller, 2001; Information Society Technologies, 2006). However, most of them can be assigned to one of these two groups.

ISA is a method that requires a small device to be used by the operator. It has five buttons that correspond to different MWL levels. The operator is asked to press a proper button with a small periodicity (usually up to several minutes). Such requirements are similar to those from the performance-based category. Thus, ISA and similar on-line methods (like PUMA and others) were rejected. The "Bedford" MWL scale was developed on the basis of the Cooper-Harper scale. It is a simple questionnaire that can be even learned by heart. It provides ten options that differ by MWL level and can be applied with the same frequency as ISA. For instance, Marinescu et al. (2017) applied an interval of 45 seconds. It could be helpful, but data received in this way requires further processing after being extracted. It is not clear how to calculate a single MWL level during the work process if the same number of aircraft gets different MWL levels.

Some methods like AHP or ACT-R (Park et al., 2018) differ from the above two patterns, but they have their

disadvantages. For instance, AHP requires a pairwise comparison of each set of conditions and multiple sub-levels of the hierarchy to be arranged. However, the binary comparison of two aircraft numbers hardly provides different results for different operators. Moreover, the comparison in AHP is required for qualitatively different workloads, while our task holds rather quantitative differences. The only method that might be considered as directly applicable to construct the continuous MWL profile is the MACE solution. It has two drawbacks, though. The first one is that all results are rather relative than absolute values. Also, it is not clear whether the same relative values of different answers can be converted to the same absolute values. The second one is that a respondent gives an answer considering only two referential values at the same time. Taking into account the above, it was decided to create a new survey form.

2.2. Survey development

The first thing to decide was whether to stick to a unidimensional or multidimensional approach (Stanton, 2005). The aim was to perform a quick and least intrusive survey efficiently and simultaneously applicable to many operators. Thus, creating a time-consuming approach like NASA-TLX, SWAT, or Workload Profile (Valdehita et al., 2004) was considered unnecessary. The generalized nature of an expected MWL measurement result provided by unidimensional methods is not necessarily a disadvantage (Tsang & Vidulich, 2012; Jordan, 1992). Received results are proven to correlate with physiological measures (Miller, 2001; Tattersall & Foord, 1996; Hamilton & Landry, 1993), be more sensitive (De Waard, 1996), have better performance (Miller, 2001). Furthermore, unidimensional methods were successfully used for multidimensional measures verification (Tsang & Velazquez, 1996) and even applied to multidimensional measurements (De Waard, 1996).

During axes selection for the horizontal one there were no options but the number of aircraft starting with 0 and ending up with 30. This number was fairly considered a sufficient cap for inexperienced ATC students. The vertical axis choice was more complicated. Most of the methods mentioned above use the scale starting from 0 and up to different top values. However, there is at least one scale used by MACE with a negative inferior limit equal to -100 . It was used its range ($-100, 100$) but changed the step value from the original 25 down to five per mark. There are several other arguments in favor of such a choice. Despite the non-linear human nature of thinking and cognition, simply asking about the MWL rate to the aircraft number could provide us with linear answers. Moreover, it wouldn't allow searching for such important parameters as MWL with maximum capacity (Goillau & Kelly, 1997), optimal performance, point of overload (Stanton, 2005; Weeda & Zeilstra, 2013; De Waard, 1996), boredom point (Hopkin, 1995), predicting performance (Wickens, 2008), etc. To avoid possible linearity in answers, we decided to formulate the survey questions in terms of boredom during under-

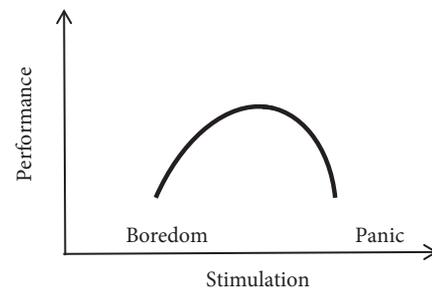


Figure 1. Generalized plot of the Yerkes-Dodson law

load, job satisfaction at the optimal MWL level, and strain caused by overload (Stanton, 2005; Hopkin, 1995). These three factors significantly influence performance (Tsang & Vidulich, 2012), which, for its part, witnesses the workload level opinion (Wang et al., 2020). The answers were expected to be in “inverted-U” form (De Waard, 1996), thus corresponding to the Yerkes-Dodson law (Figure 1) (Teigen, 1994; Rusnock & Borghetti, 2018).

Negative values foreseen for performance self-assessment were expected to appear in case of boredom and arousal (De Waard, 1996); while the closer MWL is to the optimal level, the higher self-esteem values should become (Hopkin, 1995). It seems natural for an ATC operator since he is satisfied with his job when he can use his professional skills and knowledge in the best way.

The final part is the question formulation. There was one question given to the operators after grid explanation. The performance of the ATC operator is at the highest level when he supports the maximum number of aircraft with no significant overload, i.e. when he fits the task in the best way. In contrast, the performance is worst when no workload is given or its value is overwhelming. In other words, this happens when his utility as a professional is at the highest possible value. Here utility unites two margin states when the operator is heavily underloaded or overloaded and, thus, can't act perfectly. Therefore, the question was formulated as follows: “Imagine that you get no satisfaction/demonstrate worse utility (-100 value) of your professional activity at 0 aircraft point (because of boredom and lack of MWL) and at 30 aircraft point (because of overwhelming stress). Please plot the graph of your satisfaction level across the given range. Pay special attention to the point of best performance/highest utility ($+100$ value)”.

Taking into account the satisfaction component in the utility self-estimation. The range to also consider the operator's estimate of utility as an estimate of his comfort level was used but the step value was changed. Here having no workload brings discomfort by making him useless (initial negative utility). As workload rises, the comfort level increases (utility rises, as well). The operator's estimate passes through the 0-utility point (where comfort changes a sign) and goes further. At a maximal utility point, the operator fits the task in the best way, and he understands it (having the best comfort level). Then, as the workload

becomes greater, the comfort level falls (with utility decrease). When the operator can't support sufficient safety level, his utility becomes negative again (experiencing another comfort sign change). Finally, utility falls down to zero when operator understands that he is useless, with maximal discomfort as a consequence.

Participants' experience plays no specific role, but it was decided to demand at least 150 hours of training from them. Such a requirement is caused by concerns that participants might forget their attitude after some time. This statement was considered as partially true. However, the tasks of ATC students are very similar, the variety of scenarios is low, and the questions were not about particular situation but rather about the overall experience, and, finally, students perform these tasks several times a week. All this allows ignoring the opinion forgetting by simply applying certain simulation training time threshold.

To sum up, in the proposed method, MWL is measured as a self-estimated operator's opinion on his utility during the task performance. The received grid offers a continuous (taking into account that it's impossible to measure single aircraft's portions) esteem and has referential points of -100 and $+100$. The chart requires about 5 minutes to be drawn on the grid with minor further data processing.

2.3. Strengths and weaknesses

Summarizing all of the above and based on sections 2.1 and 2.2, the strengths and weaknesses of the proposed method can be listed in comparison to the existing solutions in Table 1.

2.4. Expected benefits

Due to its properties, the method is expected to support solving such important problems as participants' comparison (Stanton, 2005), reference point absence, different workload definition, tasks similarity (Wickens, 2002), quantitative MWL definition (Wang et al., 2020), measurement continuity (Miller, 2001), metric sensitivity (Bernhardt et al., 2019), the "Multiple resource the-

ory" further research (Bommer, 2016), dynamic MWL assessment (Iqbal et al., 2020; Kostenko et al., 2016), and workload redline determination (De Waard, 1996). It also contributes to continuous MWL profiles' construction as an alternative to analytical methods (Rusnock & Borghetti, 2018) and should be tested for such an urgent topic as MWL prediction (Bommer, 2016; Kirwan et al., 1997; Vidulich et al., 1991; Wickens, 2002). While addressing important issues and challenges listed in a survey by Young et al. (2015), it should be also emphasized the possible contribution in both underload and overload redlines determination.

Perhaps the last benefit of the proposed method worth mentioning is the overall high level of acceptance of subjective methods among operators (Rehmann, 1995). It is one of seven important MWL measurement methods properties (Valdehita et al., 2004). Recalling the other six, we may assume the following. The sensitivity of subjective methods is also considered high. Diagnosticity is expected to be low because of the method's nature. Selectivity and validity should be high, supported by the procedural conditions. Intrusiveness is absent as the survey is not performed during the work process. Implementation requirements are also low. The last property is reliability, and its value to be tested and verified during the method exploitation.

3. Case study

3.1. General circumstances

The survey was performed in NAU, Kyiv, Ukraine, and KFA, Kropyvnytskyi, Ukraine. A total of 132 students of the specialty "Air traffic management" were involved. Participants were at the 4th and 5th grades (30 from NAU, 4th grade; 78 from KFA, 4th grade; 24 from KFA, 5th grade). Of all those participating in the survey, 125 students answered this part (26 from NAU, 4th grade; 75 from KFA, 4th grade, all from KFA, 5th grade). Out of 125 respondents, 1 student answered fundamentally wrong (KFA, 4th grade). His answer was completely deleted. This leaves 124 accepted answers (26 from NAU, 4th grade; 74 from KFA, 4th grade, 24 from KFA, 5th grade).

Table 1. Comparison of methods

The method(s) to be compared with the existing MWL methods	Comparative strengths of the new proposal	Comparative weaknesses of the new proposal
NASA-TLX, SWAT, Workload Profile, Bommer framework, etc.	Less time consuming, requires no special skills to accomplish the survey and process the results, allows constructing a continuous profile with no additional data processing	Unidimensional, thus less distinguishing; the questions and the task scale might be confusing for the participant
ISA, Bedford MWL scale, PUMA, etc.	Requires no special hardware, has higher precision, allows constructing a continuous profile with no additional data processing	Provides information after some time with the risk of its change
AHP	Less time consuming, works with quantitative values, simple to understand	Less detailed, provides somewhat superficial results
MACE	Provides absolute values with boundary values, requires no additional data processing	Requires more information to be taken into consideration while answering the questions

3.2. Participants experience

Every participant had a previous experience of at least 150 hours within the ATM simulation training and no real ATM experience. All students mastered the same curriculum. They passed various training scenarios at the same airport model (however, the scenarios were different for different institutions). Training sessions took place for two years (3rd and 4th grades) in compliance with the education program. Students of the 5th grade had no training experience for about 8 months by the time of the survey with an overall training experience of 200 hours.

3.3. The grid

The ATC students' task was to specify their self-assessed utility for each number of aircraft within the grid shown in Figure 2.

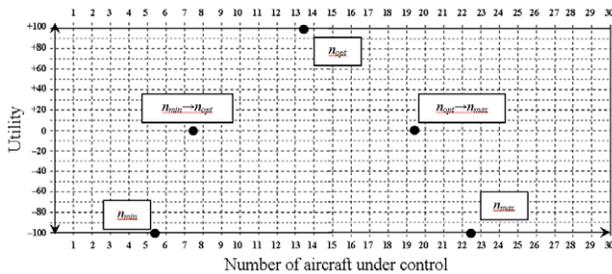


Figure 2. An example of the workload grid with key points

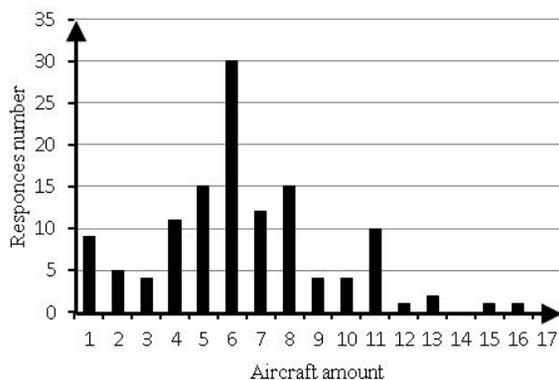
The respondents were asked to select the following key points on the grid: n_{min} with the utility of such workload equal to $u(n_{min}) = -100$; n_{opt} with the utility of such workload equal to $u(n_{opt}) = +100$; n_{max} with the utility of such workload equal to $u(n_{max}) = -100$.

Naturally, the interval $[n_{min}, n_{opt}]$ with conditions switching from boredom to the optimal workload has increasing value of the utility function. That's why respondents were asked to specify the transition point $n_{min \rightarrow opt}$ with utility equal to $u(n_{opt \rightarrow max})$. At the same time, on the interval with the rising workload $[n_{opt}, n_{max}]$, the utility function decreases. Respondents were asked to plot another transition point $n_{opt \rightarrow max}$ with utility equal to $u(n_{opt \rightarrow max})$. Utility change from -100 up to $+100$ and vice versa is considered the expected response pattern.

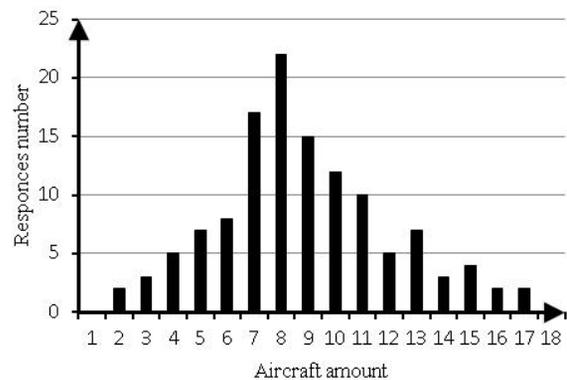
4. Data processing

Five key points of the workload utility estimation functions were analyzed. Their distributions were calculated. The corresponding plots are presented in Figure 3. For n_{min} data, the peak point is 6 aircraft, for $n_{min \rightarrow opt}$ data, the peak point is 8 aircraft. The responses distribution for n_{opt} shows bimodal results with the peak values of 10 and 15 aircraft. Data on $n_{opt \rightarrow max}$ and n_{max} has no distinct peak points.

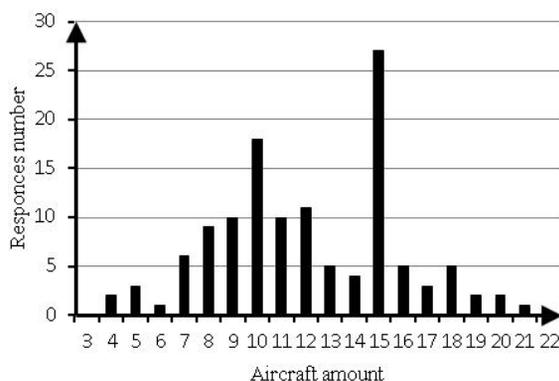
Five key points' statistics for all students and the 4th and 5th grades' students separately are presented in Tables 2–4.



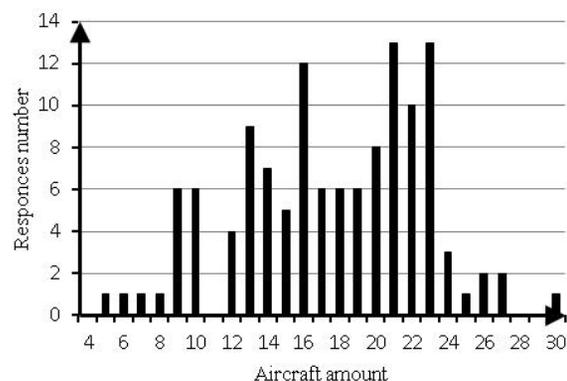
a)



b)



c)



d)

Figure 3. To be continued

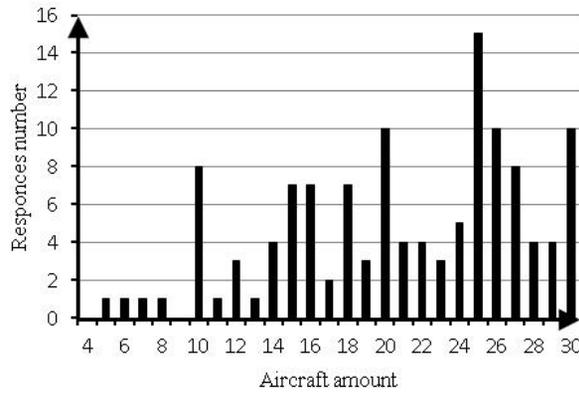


Figure 3. Key points' values distribution for the ATC students workload utility function: a - n_{min} ; b - $n_{min \rightarrow opt}$; c - n_{opt} ; d - $n_{opt \rightarrow max}$; e - n_{max}

Table 2. Statistics of the full sample key points

Key point	Ex	Var	SD	Sk	Kr
n_{min}	6.362	9.111	3.018	0.45	0.511
$n_{min \rightarrow opt}$	8.830	10.206	3.194	0.375	0.019
n_{opt}	12.104	13.671	3.697	0.071	-0.562
$n_{opt \rightarrow max}$	17.572	25.954	5.094	-0.274	-0.536
n_{max}	20.935	41.15	6.415	-0.472	-0.721

Table 3. Statistics of the 4th grade sample key points

Key point	Ex	Var	SD	Sk	Kr
n_{min}	6.45	8.088	2.844	0.58	0.747
$n_{min \rightarrow opt}$	8.91	10.203	3.194	0.44	0.0658
n_{opt}	11.93	14.409	3.796	0.087	-0.697
$n_{opt \rightarrow max}$	17	26.767	5.174	-0.244	-0.853
n_{max}	20.26	42.8	6.542	-0.377	-0.86

Table 4. Statistics of the 5th grade sample key points

Key point	Ex	Var	SD	Sk	Kr
n_{min}	6	13.739	3.706	0.313	-0.081
$n_{min \rightarrow opt}$	8.5	10.528	3.244	0.125	-0.073
n_{opt}	12.833	10.406	3.226	0.265	0.427
$n_{opt \rightarrow max}$	19.958	16.216	4.027	0.347	0.841
n_{max}	23.75	25.587	5.058	-0.627	-0.324

The mean can be used, because its values are close to median and don't significantly deviate from mode for all key points (Table 5).

Table 5. Mean, mode and median values for key points

Key point	Mean	Median	Mode
n_{min}	6.362903	6	6
$n_{min \rightarrow opt}$	8.830645	8	8
n_{opt}	12.10484	12	15
$n_{opt \rightarrow max}$	17.57258	18	23
n_{max}	20.93548	22	25

Twenty-five respondents indicated several optimal workload values (plateaus) instead of a single one ($\approx 20\%$ of all participants). Figure 4 shows plateaus lengths provided by these respondents. More than half of them (56%) claim two optimal workload values as perfectly acceptable. The plateau with length of one corresponds to two values and represents two numbers of aircraft under control with maximal values of self-assessed performance (e.g. 7 and 8 aircraft).

The sample distribution by grades and educational institutions (Figure 5) shows that the majority of double peak supporters belong to the 4th grade (left and middle columns).

Meanwhile, the 5th-grade results (right columns) are spread to the right (i.e. demonstrating wider peaks). Having 25 respondents from NAU (all of them are of the 4th grade), there were only two people with the plateau of maximal work performance observed. Such portion is equal to 8%. By contrast, the KFA 4th grade has this ratio two and a half times higher and is equal to 20%.

Seventeen respondents (13.7% of the total number) gave answers with deviations from the expected utility pattern ($-100/+100/-100$). Of those 17, there are 8 who designated the maximal work process utility value equal to 0 (6% of the total respondents).

It is important to emphasize that all participants with pattern deviation belong to three different academic groups of KFA. Also, the ratios of such respondents are

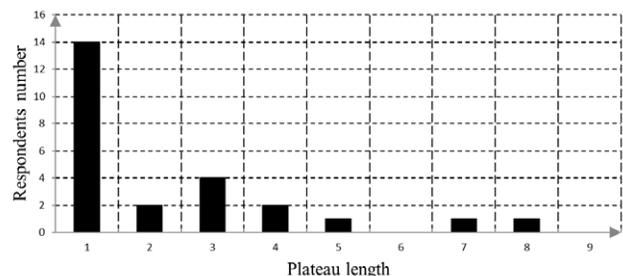


Figure 4. Distribution of the ATC students' optimal workload plateaus lengths

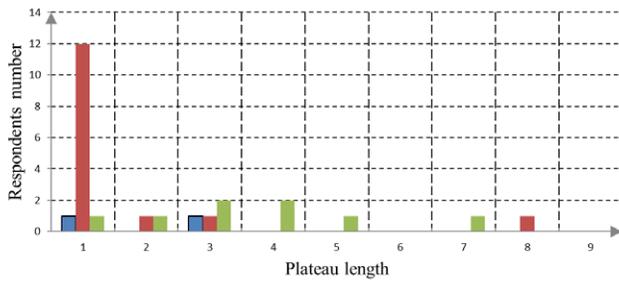


Figure 5. The number of respondents who answered with peak plateaus; left columns are for the 4th grade of NAU, the middle columns – for the 4th grade of KFA, the right right columns – for the 5th grade of KFA

close: for the first KFA group, $P_{K1} = 16.6\%$, for the second KFA group, $P_{K2} = 15.5\%$, and for the third KFA group, $P_{K3} = 19.3\%$. The general portion of deviating responses is $P_K = 17\%$.

Fit tests of five key points' goodness were carried out. For peak utility plateaus, the initial points of each plateau were processed. Three types of H_0 hypotheses were used: data distribution is normal; data distribution is exponential; data distribution is uniform. The alternative hypothesis H_1 is always directly opposite to H_0 . For normal and exponential distributions testing, the χ^2 method is used. The uniform distribution testing was performed with the help of the Cheng-Spirling method. Hypotheses testing showed the following results:

1. Not a single key points' distribution is significant given the exponential hypothesis;
2. The results for n_{min} are not significant given any distribution (normal, exponential, uniform);
3. For the full sample, the $n_{min \rightarrow opt}$, n_{opt} and $n_{opt \rightarrow max}$ results are significant given normal distribution (significance levels are $\alpha = 0.05$, $\alpha = 0.05$, $\alpha = 0.05$, p -values are .012, .068, .093);
4. For the 4th grade, the $n_{min \rightarrow opt}$, n_{opt} and $n_{opt \rightarrow max}$ results are not significant given any probability distribution (normal, exponential, uniform);
5. For the 5th grade, the $n_{min \rightarrow opt}$, n_{opt} and $n_{opt \rightarrow max}$ results are significant given normal distribution (significance level is $\alpha = 0.1$, p -values are .37, .23, .66);
6. For the full sample, the n_{max} results are significant given uniform distribution with $\alpha = 0.01$;
7. For the 4th grade, the n_{max} results are significant given uniform distribution with $\alpha = 0.01$;
8. For the 5th grade, n_{max} is significant given both normal and uniform probability distribution with the significance levels of $\alpha = 0.1$ (p -value for normal distribution is .5).

The results for normal and uniform distribution testing are presented in Table 6 (NRS correspond to no significant test result being found, "N" stands for normal distribution, "U" stands for uniform distribution).

Table 6. Significance levels for hypothesis testing results

Key point	Full sample		4 th grade		5 th grade	
	N	U	N	U	N	U
n_{min}	NRS	NRS	NRS	NRS	NRS	NRS
$n_{min \rightarrow opt}$	0.01	NRS	NRS	NRS	0.1	NRS
n_{opt}	0.05	NRS	NRS	NRS	0.1	NRS
$n_{opt \rightarrow max}$	0.05	NRS	NRS	NRS	0.1	NRS
n_{max}	NRS	0.01	NRS	0.01	0.1	0.1

5. Discussion

To get referential values without applying another MWL measurement method, the following options were chosen:

1. "Miller's magic number" (Cowan et al., 2007);
2. Regulatory documents concerning operators' work strain used in countries with similar legislation (Ukraine, Russian Federation, Belarus);
3. Maximal number of aircraft simultaneously under control used in syllabuses (equals to 16).

For the n_{min} point, $Ex_{n_{min}} = 6.36$. Miller's "magic number" which defines operational memory capacity lies in the $7 \pm 2 = [5, 9]$ range. Thus, the "magic number" contains survey results (with left shift). It witnesses boredom decrease linked to operational memory load increase in the generalized respondents' opinion.

The bimodal nature of n_{opt} data causes some confusion. However, the respondents' answers analysis allows making certain assumptions on the causes. First of all, all of the respondents with $n_{opt} = 15$ were from KFA. Secondly, their distribution across the KFA groups has no significant differences given portions $P_{K1}^{n_{opt}=15} = 33.3\%$, $P_{K2}^{n_{opt}=15} = 17.7\%$, and $P_{K3}^{n_{opt}=15} = 32.2\%$. No specific regularity is found regarding the grade. It is notable that $n_{opt} = 15$ almost matches the KFA syllabus maximum MWL. There are several explanations for it: the respondents strongly rely on the maximum aircraft number they worked with; the last scenario they worked with has the same number of aircraft; these respondents for some reason are familiar with scenarios' details.

For the n_{opt} point, $Ex_{n_{opt}} = 12.1$. This value belongs to the highly stressful level and is considered a "harmful" work conditions class according to operators' workload classification from official governmental orders. Such classification deals with the number of technological objects under simultaneous observation where the number of the objects is in range 11–25. Work conditions class: harmful, stressful work (rate 3.1).

For the n_{max} point, $Ex_{n_{max}} = 20.9$. This result also corresponds to the rate 3.1 of work condition class.

Differences between pairs of neighboring key points are presented in Table 7. According to it, an overall regularity is such that utility changes more rapidly when we observe the negative half-plane. Simultaneously, higher grade students show a wider overall range of utility change.

Table 7. The expected values' differences for the key points

Key point	Full sample	4 th grade	5 th grade
$n_{\min \rightarrow opt} - n_{\min}$	2.468	2.46	2.5
$n_{opt} - n_{\min \rightarrow opt}$	3.27	3.02	4.33
$n_{opt \rightarrow \max} - n_{\max}$	5.468	5.07	7.12
$n_{\max} - n_{opt \rightarrow \max}$	3.363	3.26	3.8

Since the $n_{\min \rightarrow opt}$ and $n_{opt \rightarrow \max}$ key points separate the negative utility from the positive one, they can be considered two redlines separator candidates. A trivial pair of such candidates is n_{\min} and n_{\max} . For now, it is difficult to say which particular value should be taken for both ranges. It can be one of the boarder points, as well as any other calculated between them. However, it can be concluded that regarding the particular definition, redlines are situated in the ranges $(n_{\min}; n_{\min \rightarrow opt})$ and $(n_{\max \rightarrow opt}; n_{\max})$, correspondingly.

$$\begin{cases} Ex_{n_{\min,5}} < Ex_{n_{\min}} < Ex_{n_{\min,4}} \\ Ex_{n_{\min \rightarrow opt,5}} < Ex_{n_{\min \rightarrow opt}} < Ex_{n_{\min \rightarrow opt,4}} \\ Ex_{n_{opt,4}} < Ex_{n_{opt}} < Ex_{n_{opt,5}} \\ Ex_{n_{opt \rightarrow \max,4}} < Ex_{n_{opt \rightarrow \max}} < Ex_{n_{opt \rightarrow \max,5}} \\ Ex_{n_{\max,4}} < Ex_{n_{\max}} < Ex_{n_{\max,5}} \end{cases} \quad (1)$$

The comparison of different grades' results shows that the 4th-grade students have their opinions more concentrated comparing to the 5th-grade students. It is proven by the expected values' shift towards less aircraft for the 4th-grade students and towards more aircraft for the 5th-grade students on the proposed range. The box plot of key values' also supports this conclusion (Figure 6).

Statistics' skew decreases from low to high workload. Calculated skew coefficients witness that in respondents' opinion, the key point distribution tends to the "central" point, i.e. to the point of optimal workload.

$$Sk_{n_{\min}} > Sk_{n_{\min \rightarrow opt}} > Sk_{n_{opt}} > Sk_{n_{opt \rightarrow \max}} > Sk_{n_{\max}} \quad (2)$$

This is true for the full sample and the 4th-grade responses. The 5th-grade data somewhat disturbs this tendency that can be explained by a certain lack of training.

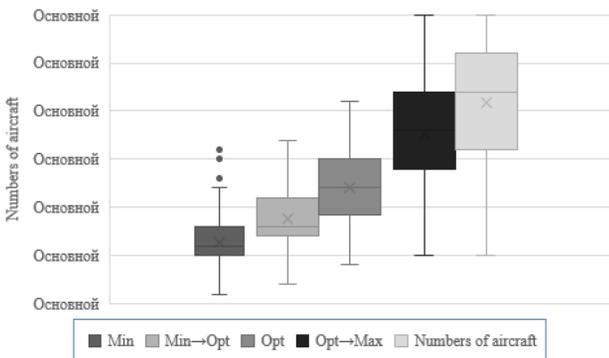


Figure 6. Key values box plot

The least skew is shown by the optimal workload point n_{opt} , where $Sk_{n_{opt}} = 0.0706$.

As per the classic formula, the kurtosis is calculated with even power and thus, cannot be of negative values. However, many literature sources use formulas with relative kurtosis values compared to the normal distribution. Such a presentation has its advantages. When comparing the received value with a 0 value, it can be concluded about the peak form immediately, without additional calculation. The kurtosis shift from the positive to negative values supports the assumption based on the charts presented in Figure 3. As workload increases, the opinions consistency decreases. It is also approved by the variance/standard deviation growth. The kurtosis distribution statistics by grades shows similar tendencies as for skew. The 5th-grade responses seem to be less arranged and subjected to certain distribution.

Of all 25 answers with plateau:

- 17 answers are of the 4th-grade students, that is 17% of their overall number;
- 8 answers are of the 5th-grade students, that is 33% of their overall number.

Such proportion witnesses that the 5th-grade students possess greater ATC experience (even if it is only training) and consider themselves having a high level of stable and reliable work performance, thus claiming several peak workload values (i.e. workload plateau).

The potential explanation for institution prevalence (2 from the NAU 4th-grade vs. 15 from the KFA 4th-grade) requires additional research. It can be explained by the fact that NAU is a university where the same department can teach students from different majors. At the same time, KFA is a specialized educational institution, where students experience greater immersion in the ATM environment.

A certain part of respondents demonstrated a distinct risk avoidance attitude of different strengths (13.7%) with a deviation from the $(-100/+100/-100)$ pattern. All of them are KFA students. Such answers point out an overall low motivation and a lack of professional activity orientation. This statement is supported by the determined values of the main decision-making dominants: all considered participants show risk inclination, i.e. a tendency for failure avoidance (Reva et al., 2016; Borsuk, 2017).

The goodness of fit results implies that the distribution of minimal workload key point is undetermined. This makes sense as there is no strain and pressure on ATCs at this MWL level. However, in the boredom research aspect, this result means that the overall distribution law is yet to be found for the population.

In contradiction, the workload levels presented with the $n_{\min \rightarrow opt}$, n_{opt} and $n_{opt \rightarrow \max}$ points can be treated as if the population corresponds to the normal distribution. The notable part is that the sample of the 4th grade is less consistent than the full sample. It is caused by the solid and confident opinions of the 5th-grade students after their learning. Double distribution for n_{\max} that satisfies both uniform and normal distribution for the 5th grade requires further research.

Based on the key points' analysis, less arranged skew and kurtosis, greater variance/standard deviation for the 5th grade, wider [n_{\min} , n_{\max}] range, greater key points' expected values differences, greater plateaus lengths, it can be stated that the 4th-grade opinions are more consistent on the pattern change while the 5th-grade opinions are more consistent on each particular point distribution and prove higher confidence.

Pronounced normal distribution tests' results witness that in general, the 5th-grade students' opinions on each proposed key point can be processed with analytical methods like Analysis of variance (ANOVA). This is not true for the 4th grade.

Most of the respondents, who demonstrated workload plateau, all of the respondents with pattern deviations, and all of the respondents, who gave the $n_{opt} = 15$ answer belong to KFA. The number of deviations in these answers can't be explained with just a low number of respondents. Thus, it is possible that there is some fundamental difference in the educational process between these two institutions. The most probable explanation could be the difference in the institutions' focus, instructors, particular scenarios. However, more data should be collected and processed to discover whether any of the educational processes are better.

Finally, after using the method, it is possible to conclude on its properties. It was easily accepted by the audience; sensitivity was high not only in terms of particular operators' differentiation but in the sense of operators' groups' distinction as well. Diagnosticity appeared to be arguable due to differences in the key points' expected values. It's not only about inequality for the upper (positive utility) and lower (negative utility) half-planes, but for the left (boredom avoidance) and right (strain increase) parts. Selectivity, intrusiveness, and implementation requirements have no change as to what was predicted. The combination of all these properties allows the method to surpass alternatives in the resolved problem. Unfortunately, reliability (comparison to the alternatives precision) was not tested during this study and still to be evaluated. The method limitations list includes the applied single dimension, possible confusion with the question and the answers, the time delay between the latest work experience and opinion expression.

The proposed method can be used for the participants' comparison across the proposed MWL range. Also, the key points' statistics applies to groups' comparison. Participants have to recollect their overall attitude towards different workload levels at the same time while completing the survey. Also, they imagine three points of the most negative and the most positive utility. These three points can be taken as referential values for comparisons. The problem of different workload definition is compensated with the help of the question formulation where workload measurements are expressed with own utility. Due to the simple controllable object definition, the quantification of MWL is possible. Each particular result represents a continuous index set that allows defining the redline and as a matter of fact, can be considered as the operator's profile.

According to the sample data, there are two ranges of potential minimal and maximal redlines. Their particular values should be tested taking into account cross-cultural features. When considering the most probable optimal workload plateau range, the $x = 12 \pm 1$ number of aircraft is recommended as a minimal satisfying number of aircraft under control for training purposes.

Further data processing of the $n_{\min \rightarrow opt}$, n_{opt} , $n_{opt \rightarrow max}$ and n_{max} key points should take into account their normal distribution. It allows applying methods like ANOVA, which relies on the distribution normality. More studies are required on the n_{\min} key point.

Conclusions

The proposed method is a simple alternative to the existing methods of MWL measurement. As it comes from the literature review, it is the second subjective unidimensional method straightly applicable to create an MWL profile. These profiles provide clearly interpreted data on the absolute scale. The MWL measurement with the operator's opinion on utility during the task performance provides continuous data with referential points. It can be used in a distant way, which is vital in the light of COVID-19 worldwide restrictions. It requires no special hardware, only minor data processing, and is easily accepted by the audience.

It is proven that the more specialized the educational institution is, the better students' self-assessment of capabilities occurs after their learning. It leads to both disappointment and self-overestimation of students concerning flight safety support depending on a particular pupil.

The further ATCs students' workload research should be performed using a set of surveys with the subsequent comparison of their results for different samples, additional analysis of respondents' results with performance plateau and pattern deviation, cross-verification analysis of the received results with same respondents' risk attitude, aspiration levels dedication based on the received results, synthesis of special combined criteria intended to verify the stability in higher educational institutions' educational process, application of the ANOVA method for different student's groups, and performing the same survey with real ATCs and comparing the results. The reliability of the proposed method should be tested, as well. More distinguishing factors like age, gender, instructor, country of education, experience level, and hardware should also be taken into account in further researches.

References

- Bernhardt, K., Poltavski, D., Petros, T., Ferraro, F., Jorgenson, T., Carlson, C., Drechsel, P., & Iseminger, C. (2019). The effects of dynamic workload and experience on commercially available EEG cognitive state metrics in a high-fidelity air traffic control environment. *Applied Ergonomics*, 77, 83–91. <https://doi.org/10.1016/j.apergo.2019.01.008>
- Bommer, S. (2016). *A theoretical framework for evaluating mental workload resources in Human Systems design for manufacturing operations*. Wright State University, Dayton, Ohio, US.

- Borsuk, S. (2017). New methods for air traffic controller main solution taking dominant determination concerning their attitude to risk. *Logistics and Transport*, 1(33), 25–29.
- Braarud, P. Ø. (2020). An efficient screening technique for acceptable mental workload based on the NASA Task Load Index – development and application to control room validation. *International Journal of Industrial Ergonomics*, 76, 102904. <https://doi.org/10.1016/j.ergon.2019.102904>
- Chiappe, D., Morgan, C. A., Kraut, J., Ziccardi, J., Sturre, L., Strybel, T. Z., & Vu, K.-P. L. (2016). Evaluating probe techniques and a situated theory of situation awareness. *Journal of Experimental Psychology: Applied*, 22(4), 436–454. <https://doi.org/10.1037/xap0000097>
- Cowan, N., Morey, C. C., & Chen, Z. (2007). The legend of the Magical Number Seven. In S. Della Sala (Ed.), *Tall tales about the mind and brain: Separating fact from fiction* (pp. 45–59). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198568773.003.0005>
- De Waard, D. (1996). *The measurement of drivers' mental workload*. University of Groningen.
- Efthymiou, M., Fichert, F., & Lantzsch, O. (2019). Workload perception of air traffic control officers and pilots during continuous descent operations approach procedures. *Aviation Psychology and Applied Human Factors*, 9(1), 2–11. <https://doi.org/10.1027/2192-0923/a000154>
- Gao, Q., Wang, Y., Song, F., Li, Z., & Dong, X. (2013). Mental workload measurement for emergency operating procedures in digital nuclear power plants. *Ergonomics*, 56(7), 1070–1085. <https://doi.org/10.1080/00140139.2013.790483>
- Goillau, P. J. & Kelly, C. J. (1997). MALvern Capacity Estimate (MACE) – A proposed cognitive measure for complex systems. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics, Vol. 1: Transportation Systems* (pp. 219–225). Routledge. <https://doi.org/10.4324/9781315094496-22>
- Hamilton, K. M., & Landry, L. N. (1993). Measuring subjective workload: When is one scale better than many? *Human Factors*, 35, 579–601. <https://doi.org/10.1177/001872089303500401>
- Hopkin, V. D. (1995). *Human factors in air traffic control*. Taylor & Francis.
- Information Society Technologies (IST). (2006). *Adaptive integrated Driver-vehicle interface. Subjective assessment methods for workload*. http://www.aide-eu.org/pdf/sp2_deliv_new/aide_d2_2_6.pdf
- Iqbal, M. U., Babji, S., & Rajagopalan, S. (2020). Dynamic assessment of control room operator's cognitive workload using Electroencephalography (EEG). *Computers & Chemical Engineering*, 141, 106726. <https://doi.org/10.1016/j.compchemeng.2020.106726>
- Jordan, C. S. (1992). *Experimental study of the effect of an instantaneous self assessment workload recorder on task performance. DRA Technical Memorandum (CAD5)*. DRA Maritime Command Control Division.
- Kirwan, B., Evans, A., Donohoe, L., Kilner, A., Lamoureux, T., Atkinson, T., & MacKendrick, H. (1997). Human factors in the ATM system design life cycle. In *FAA/Eurocontrol ATM R&D Seminar* (pp. 16–20). Paris, France.
- Kostenko, A., Rauffet, P., Chauvin, C., & Coppin, G. (2016). A dynamic closed-looped and multidimensional model for Mental Workload evaluation. *IFAC-PapersOnLine*, 49(19), 549–554. <https://doi.org/10.1016/j.ifacol.2016.10.621>
- Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science*, 42(4), 237–270. [https://doi.org/10.1016/S0925-7535\(03\)00047-X](https://doi.org/10.1016/S0925-7535(03)00047-X)
- Marinescu, A. C., Sharples, S., Ritchie, A. C., López, T. S., McDowell, M., & Morvan, H. P. (2017). Physiological parameter response to variation of mental workload. *Human Factors*, 60(1), 31–56. <https://doi.org/10.1177/0018720817733101>
- Miller, S. (2001). *Literature review. Workload measures*. National Advanced Driving Simulator, The University of Iowa.
- Nachreiner, F. (1995). Standards for ergonomics principles relating to the design of work systems and to mental workload. *Applied Ergonomics*, 26(4), 259–263. [https://doi.org/10.1016/0003-6870\(95\)00029-C](https://doi.org/10.1016/0003-6870(95)00029-C)
- Park, S., Jeong, S., & Myung, R. (2018). Modeling of multiple sources of workload and time pressure effect with ACT-R. *International Journal of Industrial Ergonomics*, 63, 37–48. <https://doi.org/10.1016/j.ergon.2017.07.003>
- Rahman, S. M. B. A., Borst, C., van Paassen, M. M., & Mulder, M. (2016). Cross-sector transferability of metrics for air traffic controller workload. *IFAC-PapersOnLine*, 49(19), 313–318. <https://doi.org/10.1016/j.ifacol.2016.10.561>
- Rehmann, A. J. (1995). *Handbook of human performance measures and crew requirements for Flightdeck research*. Report. Bureau of transportation statistics, U.S. Department of Transportation. <https://doi.org/10.1037/e664922007-001>
- Reva, O. M., Borsuk, S. P., Bala, M. M., & Peyman, M. S. (2016). New approach to determination of main solution taking dominant of air traffic controller during flight level norms violation. In F. Rebelo & M. Soares (Eds.), *Proceedings of the AHFE 2016 International Conference on Human Factors in Transportation* (pp. 137–147). Springer. https://doi.org/10.1007/978-3-319-41682-3_12
- Roscoe, A. H., & Ellis, G. A. (1990). *A subjective rating scale of assessing pilot workload in flight: A decade of practical use, a technical report*. Procurement Executive, Ministry of Defence Farnborough, Hampshire.
- Rusnock, C. F., & Borghetti, B. J. (2018). Workload profiles: A continuous measure of mental workload. *International Journal of Industrial Ergonomics*, 63, 49–64. <https://doi.org/10.1016/j.ergon.2016.09.003>
- Socha, V., Hanáková, L., Valenta, V., Socha, L., Ábela, R., Kušmírek, S., Pilmannová, T., & Tecl, J. (2020). Workload assessment of air traffic controllers. *Transportation Research Procedia*, 51, 243–251. <https://doi.org/10.1016/j.trpro.2020.11.027>
- Stanton, N. A. (2005). *Handbook of human factors and ergonomics methods*. CRC Press. <https://doi.org/10.1201/9780203489925>
- Tattersall, A. J., & Foord, S. (1996). An experimental evaluation of instantaneous self assessment as a measure of workload. *Ergonomics*, 39(5), 740–748. <https://doi.org/10.1080/00140139608964495>
- Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology*, 4(4), 525–547. <https://doi.org/10.1177/0959354394044004>
- Tsang, P., & Velazquez, V. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, 39(3), 358–381. <https://doi.org/10.1080/00140139608964470>
- Tsang, P. S., & Vidulich, M. A. (2012). *Mental workload and situation awareness*. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 243–268). John Wiley & Sons, Inc. <https://doi.org/10.1002/0470048204.ch9>
- Valdehita, S., Ramiro, E., García, J., & Puente, J. (2004). Evaluation of subjective mental workload: a comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology*, 53(1), 61–86. <https://doi.org/10.1111/j.1464-0597.2004.00161.x>

- Vidulich, M. A., Ward, G. F., & Schueren, J. (1991). Using the Subjective Workload Dominance (SWORD) technique for projective workload assessment. *Human Factors*, 33(6), 677–691. <https://doi.org/10.1177/001872089103300605>
- Wang, P., Fang, W., & Guo, B. (2020). A measure of mental workload during multitasking: Using performance-based Timed Petri Nets. *International Journal of Industrial Ergonomics*, 75, 102877. <https://doi.org/10.1016/j.ergon.2019.102877>
- Weeda, C., & Zeilstra, M. (2013). *Prediction of mental workload of monitoring tasks*. In N. Dadashi, A. Scott, J. R. Wilson, & A. Mills (Eds.), *Rail human factors: Supporting reliability, safety and cost reduction* (pp. 633–640). CRC Press. <https://doi.org/10.1201/b13827-83>
- Wickens, C. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science*, 3(2), 159–177. <https://doi.org/10.1080/14639220210123806>
- Wickens, C. (2008). Multiple resources and mental workload. *Human Factors*, 50(3), 449–455. <https://doi.org/10.1518/001872008X288394>
- Yan, S., Wei, Y., & Trana, C. C. (2019). Evaluation and prediction mental workload in user interface of maritime operations using eye response. *International Journal of Industrial Ergonomics*, 71, 117–127. <https://doi.org/10.1016/j.ergon.2019.03.002>
- Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, 58(1), 1–17. <https://doi.org/10.1080/00140139.2014.956151>
- Zeier, H. (1994). Workload and psychophysiological stress reactions in air traffic controllers. *Ergonomics*, 37, 525–539. <https://doi.org/10.1080/00140139408963668>

Notations

Variables and functions

- Ex* – expected value;
- Kr* – kurtosis;
- n_{\max} – number of aircraft in responsibility area that leads to impossibility of the very control;
- $n_{\min \rightarrow opt}$ – transition point with zero utility value on the range from minimal to optimal point;
- n_{\min} – number of aircraft in responsibility area which corresponds to the inferior workload margin;
- n_{opt} – number of aircraft in responsibility area that might be called optimal;
- $n_{opt \rightarrow t_{\max}}$ – transition point with zero utility value on the range from optima to maximal point;
- P_{K1} , P_{K2} , P_{K3} – the ratios of three different academic groups of KF;
- SD* – standard deviation;
- Sk* – skew;
- $u()$ – workload utility;
- Var* – variance;

Abbreviations

- ACT-R – Adaptive Control of Thought – Rationale;
- AHP – Analytic Hierarchy Process;
- ATC – air traffic control;
- ATM – air traffic management;
- ISA – Instantaneous self-assessment of workload technique;
- KFA – Kirovohrad Flight Academy;
- MWL – Mental workload;
- MACE – Malvern Capacity Estimate;
- NASA-TLX – National Aeronautics and Space Administration Task Load Index;
- NAU – National Aviation University;
- PUMA – Performance and Usability Modelling in ATM;
- SWAT – Subjective Workload Assessment Technique;
- SWORD – Subjective Workload Dominance.