

## Article

# Assessing Mental Workload in Dual STEM–Air Force Language Listening Practice

José Luis Roca-González <sup>1,\*</sup>, Juan-Antonio Vera-López <sup>2</sup> and Margarita Navarro Pérez <sup>3</sup>

<sup>1</sup> Department of Engineering and Applied Technologies, University Centre of Defence at the Spanish Air Force Academy, 30720 San Javier, Spain

<sup>2</sup> Department of Sciences, University Centre of Defence at the Spanish Air Force Academy, 30720 San Javier, Spain; juanantonio.vera@ud.upct.es

<sup>3</sup> Facultad de Educación de Ciudad Real, Universidad de Castilla la Mancha, 13001 Ciudad Real, Spain; margarita.navarro@uclm.es

\* Correspondence: jluis.roca@ud.upct.es

**Abstract:** Cognitive workload analysis is an important aspect of safety studies at the Spanish Air Force Academy where students must complete a dual academic curriculum based on military pilot training combined with an industrial engineering degree. Recently, a mental workload assessment (MWA) and forecasting model based on Shannon's law from information theory (IT) has been published; it proposes a new mathematical procedure (MWA-IT) that defines a workload index that could be extrapolated to other case studies. The aim of this study was to adapt this model to the Spanish University Centre of Defence to calculate the mental workload caused by the listening practice in English as a foreign language. In addition, a contrasting methodology, the NASA task load index (NASA-TLX), was applied to validate the proposed model using the error study provided by SMAPE and MSE. The results established an expected reference baseline for MWA-IT in English listening that is between 36 and 92 at the end of the four courses, which corresponds to the students that start with the lowest English level (higher workload = 92) and the ones with the highest English level certification (lowest workload = 36); meanwhile, the NASA-TLX result was between 49.8 and 193.7 for the same circumstances. The main difference is that MWA-IT can be predicted with 41% less deviation than can NASA-TLX and does not require the completion of a questionnaire following the activities. Finally, the study also highlights the fact that that nearly 65% of the workload was caused by the first two courses, when the advanced STEM subjects were taught and the pilot learning and practice program had not yet begun. This methodology may help the teachers in charge to redesign or add new content depending on the expected workload reference.

**Keywords:** cognitive workload; second language learning; information theory; air forces studies



**Citation:** Roca-González, J.L.; Vera-López, J.-A.; Navarro Pérez, M. Assessing Mental Workload in Dual STEM–Air Force Language Listening Practice. *Aerospace* **2024**, *11*, 147. <https://doi.org/10.3390/aerospace11020147>

Academic Editor: Konstantinos Kontis

Received: 10 November 2023

Revised: 4 December 2023

Accepted: 22 December 2023

Published: 11 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Workload analysis is often used as a measurement tool to identify the human factor requirements of a specific task under the framework of labour safety and security. Moreover, given that there are several surrounding factors that may affect critical activities during the military pilot learning process, there are some subjects that must be studied simultaneously; the learning of English as a foreign language is a clear example of this. Thus, when such a workload methodology is applied to the design and selection of the teaching contents of English as a foreign language in an industrial engineering degree combined with military studies, the paradigms shift. This is mainly because the interpretation of the European Credit Transfer and Accumulation Systems (ECTS) focuses on expected working hours rather than the complexity of achieving the specific goals of each subject, which students may perceive in different ways depending on their own competencies. The students considered here experience double pressure due to the fact that the military career and

degree take up a considerable amount of their time, in addition to the added day and night duties and constraints of a military life.

For example, the process of designing a course unit description for an English language program that is in accordance with the academic requirements for an industrial engineering degree requires a complete understanding of how students face and achieve progressive cross-curricular technical competencies while learning a foreign language skill. More specifically, the Common European Framework of Reference for Languages provides useful guidelines for students undertaking English language programs; these programs start in the first academic year of graduate studies; then, the difficulty level increases every year to enhance the oral and writing competencies of the students.

English for academic purposes (EAP) has become an indispensable part of college English education. However, the measurement of the effectiveness of English for general academic purposes has been, and continues to be, a subject worthy of study [1]. A multi-disciplinary analysis is required to achieve a better understanding of linguistic programs combined with academic competencies. Therefore, EAP teaching can turn out to be a challenge for teachers when deciding the content of the syllabus of an academic program and, more specifically, when trying to fill the gap between the target needs and the students' needs in English for academic environmental purposes (EAEP) [2]. Hence, constant feedback of information is needed during an EAP course. Through this feedback, the realisation of the competency potential of academic subjects in engineering education may lead to a reduction in workload for both students and teachers [3]. This provides a new perspective on foreign language teaching (FLT) where the optimisation of the students' workload may help them to achieve professional competence in the linguistic domain.

Past research has focused on English language teaching from the perspective of the instructors in order to determine the professional development needs [4]; however, more recent research has shifted the focus to student effort, particularly during speech understanding, by applying cognitive workload indicators to design new courses [5]. This new approach could help to reduce the cognitive workload in designing the content of each subject. Moreover, the courses on the study of English for academic purposes are commonly designed as four-year programs, where the level of difficulty is increased yearly.

Establishing the same reference baseline for all the case studies in relation to the cognitive workload is a mistake, as the capacity and quality of the cognitive resources differ for each subject; thus, for the delivery of similar levels of performance, the subjects may have different amounts of spare cognitive capacity remaining [6]. Therefore, the cognitive workload index should be used as a reference baseline for each student and case study given that each case may translate into a different self-developed mental workload scale.

Regarding the research on cognitive effort during learning, there are some physiological signals, such as heart rate variability, that denote a lack of sensitivity to potential changes in invested cognitive effort during skill acquisition as a function of the practice conditions [7]; thus, future research on experiments utilising behavioural measures (e.g., recall success) and perceived cognitive effort scales such as NASA-TLX is needed [8].

A heart rate analysis during an English listening test may indicate the anxiety factor linked to workload during the performance of such a test, but the focus required for English listening learning in an English as a foreign language program is more extended and demanding. Authors such as Alshabeb, Alsaie, and Albasheer claimed that speaking and listening skills are often not sufficiently addressed by conventional courses and that the most effective way to tackle this problem is not to alter classroom practice but to provide a different structure to the course, allowing more flexible learning that is adjusted to individual requirements [9]; this could be achieved by using a workload index as a baseline reference.

In connection with EAEP, listening as a skill is a topic of study that extends the existing knowledge of the natural spoken language into professional communication, where information management affects the workload of conversational partners and may manifest through psychotherapeutic factors [10]. For example, a study has recently been

conducted on home healthcare professionals' communication skills, where the influence of language barriers has been analysed [11] as a workload indicator affecting the human resources in charge of providing such a service.

Within the field of airline operations and management, air traffic management and risk and reliability are matters of study related to the aerospace domain. Human factors are the primary cause of incidents and accidents in the civil aviation industry. Among these factors, communication errors are the most critical [12]; such errors are mainly due to a listening-related misunderstanding. The latter highlights the relevance of English listening learning from the early stages of pilot training studies.

The case study covered in this paper focuses on students working towards a dual degree in industrial engineering and military pilot training, where English as a foreign language is an obvious and a mandatory requirement. Therefore, the aim of the research is to provide for such students a workload baseline reference when the English listening learning practice takes place.

Although the novelty of this research is its focus on the application of a workload assessment methodology based on information theory as extracted from a case study centred on pilot workload, it is necessary to summarise a review of workload measurement methodologies. To this methodology, it is necessary to add a relevant academic framework for English as a foreign language and for specific purposes to gather the existing knowledge that is to be taken into consideration.

### *1.1. The Research Proposal*

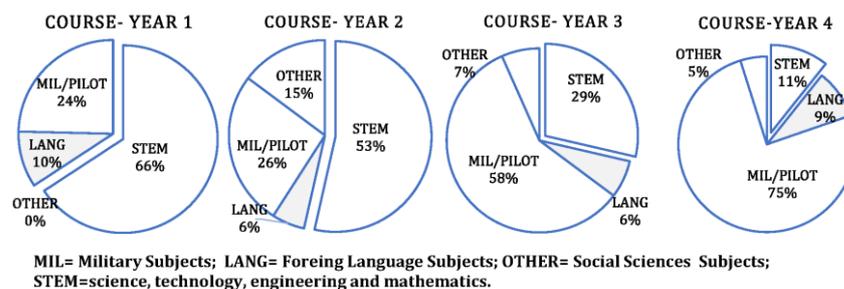
Industrial engineering study programs must consider the following EAP requirements: formal learning experience, achievement motivation, and learning needs [13]. The syllabi of English language courses should be designed in such a way that engineering students are made to feel instrumentally motivated; this can be accomplished by decreasing their cognitive workload.

This paper proposes a cognitive workload study within the scope of EAP; it focuses on listening practice in the context of the industrial engineering degree taught at the University Centre of Defence at the Spanish Air Force Academy, where the students must complete their military education and engineering studies simultaneously. The engineering curriculum is defined by 240 ECTS of the European Credit Transfer and Accumulation System, where 1 ECTS involves 10 h of face-to-face classes; up to 25–30 h of studying or autonomous work is also required. The curriculum planning is mainly built around a four-year programme, which is extended by one additional course. The latter is devoted to the writing of their end-of-degree dissertation, the final degree research report, and the professional novice practice that has a dual educational purpose (civil and military formation). In addition, the military curriculum is defined by 145.2 ECTS with the same course distribution. Here, the case study focuses on the first four academic years in order to set a time framework that is similar to that of the other degrees in industrial engineering.

The STEM subjects are algebra, physics, chemistry, technical drawing, calculus, and statistics in the first year; fluid mechanics, materials science, electrical technology, electronic instrumentation, quantitative methods, energy technology, environmental technologies, mechanical and manufacturing technology, and materials science in the second year; operation management, security and defence technology, meteorology, and communications in the third year; and project management for industrial engineering and the final degree project in the fourth/last year. The ECT percentage distribution is shown in Figure 1.

In this type of study, the framework for English language learning was defined following the transformation of the Central and East European armed forces into modern contributors to Euro-Atlantic security. For such purposes, the United Kingdom has been playing an assisting role by establishing conscript training centres to provide English language training in order to ensure the development of such academic competencies. The listening comprehension training of similar programmes offers the opportunity to analyse the workload requirements of what has been considered the most challenging skill in both

teaching and testing procedures, including listening comprehension tests, regardless of whether they are paper-based or computer-based [14]. This paper introduces an inductive methodology starting with the setting up of a working database, which was built with the information extracted from the case study records and from the statistical analysis of previously collected data. This database allowed for the calculation of a workload for listening training in English as a foreign language through the use of NASA-TLX and a method developed recently from the perspective of information theory [15] that does not require the completion of a questionnaire after performing the activity; this is significant when the activity takes place within a longer timeframe, as is the case with the process of English listening learning.



**Figure 1.** Subject typology distribution per year by percentage. Source (own).

### 1.2. Workload Measurement Methodologies

The assessment of mental workload has been extensively explored due to the escalating demands on the operators' information-processing capacities [16]. Two categories of methods, primary and secondary task measures, fall under the umbrella of performance measures for mental workload assessment. Primary task measures assume a universal human capacity for a given task, while secondary task measures consider an upper limit on the workers' ability to gather and process information; this limit prompts them to perform a secondary activity to gauge its impact on the main task.

Subjective rating measures involve the workers' perceptions of task complexity, potentially establishing a reference framework through statistical analysis. Physiological measures analyse the operators' responses to workload by examining various physiological factors. The NASA task load index (NASA-TLX), originally designed for ergonomic analysis in the aerospace industry, has gained widespread use across diverse fields due to its applicability and flexibility [17–21], which explains why it is still being used as a reference method.

The debate surrounding NASA-TLX's extensive use in cognitive workload research centres on the derivation of information from workload measures within theoretical frameworks. The NASA-TLX is considered valuable when building models for categorisation, description, explanation, and prediction, and it emphasises the importance of understanding mechanisms [22,23]. Some authors categorise mental workload methodologies into three main groups: performance, subjective, and physiological metrics. Subjective methodologies such as NASA-TLX lack clear definitions that distinguish between "effort" and "demand"; thus, clear and unambiguous language is required when defining multidimensional scales.

Although attitude and mental workload cannot be quantified directly, they can be assessed indirectly by collecting multiple measures that should be employed to comprehensively capture mental workload during task performance [24]. Comparisons between subjective workload methods such as SWAT and WP alongside NASA-TLX reveal no significant differences in intrusiveness, and they have similar sensitivity and convergent validity. NASA-TLX is recommended for the prediction of a subject's task performance due to its higher correlation with performance [25]. Other examples, such as Bedford's methodology, which uses ten scales to assess four levels, yield weak correlations between the immersive experiments and the final results, making it challenging to distinguish between different

workloads within the same range. NASA-TLX has been applied to study mental workload in conjunction with real-time physiological parameters, providing a dynamic framework for multitask workload estimation [26]. However, its use requires a thorough examination of each dimension rather than sole reliance on a global score, especially in activities such as driving, where intrinsic, extraneous, and germane loads are influenced by different variables. For example, some authors considering further dimensions of the operator's mental state have recently included mental fatigue, stress, and vigilance [27]. Others refer to a high cognitive workload status, which is defined as "a physiological state of reduced mental or physical performance capability resulting from sleep loss, extended wakefulness, circadian phase, and/or workload" (International Civil Aviation Organization).

Cognitive loads, which are defined as information-processing demands during task performance, align with cognitive theory, emphasising learning as a process of integrating new information into existing schemas. Analysing cognitive workload through immersive experiences aids in determining the best practices to minimise mental workload [28]. For instance, repeated simulation training in ureterorenoscopy in a high-fidelity setting was found to result in a continual decrease in mental workload, as the participants identified optimal strategies [29]. In this last case a model based on information theory, as an example, could help to parametrize the workload reduction as it takes into consideration the experience level of the participants.

After considering the above, we can surmise that NASA-TLX is suitable as a contrast or validation methodology when combined with the information theory that is explained in detail in the Materials and Methods Section.

### 1.3. Academic Background

NASA-TLX is widely used in academia to assess the mental and physical workload of vocational students [7]. It defines student workload expansively, encompassing the hours spent on lectures, seminars, project preparation, exams, etc. Immersive training enhances task performance by instilling careful task execution, with successful outcomes linked to the students' ability to perceive task difficulty in advance [30].

The implementation of audio cues in immersive practices, analysed through NASA-TLX, yields satisfactory results, suggesting potential benefits for pilots through the incorporation of multisensory cues, such as audio–tactile cues [31]. While NASA-TLX has been validated in numerous case studies, recent research has compared it with other scales, such as the modified Cooper–Harper (MCH) scale and the mean interbeat interval. The consensus is that MCH may be more appropriate for measuring mental workload in time-critical environments [32]. These methods, including NASA-TLX, are deemed potentially useful for multicriteria decision analysis in evaluating future human–machine complex systems or operating procedures [6].

Despite numerous workload assessment techniques, subjective ratings methods such as NASA-TLX remain the most commonly used and serve as a benchmark when comparing other measures. Information theory aids in the contrasting of expected workload models. Some of the related research explores language learning through computer games, suggesting increased cognitive demand and motivational strategies. However, there is a lack of recent publications evaluating these aspects in terms of workload requirements; thus, further research is necessary [33,34], NASA-TLX could play a relevant role in addressing these questions.

Recent research has employed NASA-TLX to measure mental workload and learners' frustration with traditional teaching procedures [35]. Previous studies highlight the importance of mental workload in understanding how teachers define and perceive course contexts and how these contexts influence instructional decisions and practices [36]. In English for academic purposes (EAP), the recent research by Bahrami, Hosseini, and Reza Atai underscores the need for specific teacher programmes to foster pedagogical content knowledge. This initiative aims to enhance the development of EAP content, making the research more relevant to the practice and contributing to the framework for improved

EAP teacher education [37]. Studying mental workload from the students' perspectives sheds light on the managerial implications of teaching English for specific purposes, which depend on the academic degree to which it is applied.

In the domain of educational technology and applied linguistics, the research emphasises the symbiotic relationship between research and teaching, suggesting that research-oriented pedagogical practices facilitate college English teachers' immersion into research practices [38].

Therefore, researching the students' workload, which is a result of the teaching practices, could enable a means of developing immersive experiences that enhance teaching capabilities and professional competencies in the academic community.

The students' mental workload indicators, which are linked to listening training units, even when collected subjectively through NASA-TLX, can inform content arrangement with a progressive increase in difficulty. This principle can extend to collaborative science learning in immersive experiences, offering valuable insights into the design of accurate simulations that foster constructive learning immersions [39]. Workload studies in immersive practices serve as a useful tool for ensuring behavioural compliance before research is conducted; assuming that the workload in the immersive experiences is primarily defined by task difficulty level and user capability [40], a workload model based on information theory that has been selected for this research can be easily introduced.

## 2. Materials and Methods

The materials needed to perform this study were extracted from teaching experiences in the industrial engineering grade at the University Centre of Defence at the Air Force Academy, and were considered together with the students' questionnaires to set up a working database. In compliance with data collection for case studies with humans as the main subjects, multiple sources of data were used, and a chain of evidence was kept. Furthermore, as the main focus of this study was to apply a recent mental workload assessment–information theory methodology, together with a contrasted NASA workload index, some missing data were gathered from assumptions based on the statistical analysis of previously collected data, as described in the Data Source Section. For this study, it was not necessary to request specific approval by the institutional review board, as the study did not require human participants. The CV workload index was calculated using random values applied to the variables according to the experiences reported by the teachers in charge. Meanwhile, NASA-TLX was calculated using the teacher satisfaction report provided by the university, for which 12 questions were used to set up a working database and which were linked afterwards with the NASA rating scale definitions.

### 2.1. Data Source

The case study information for the four-year programme of English as a foreign language for specific and academic purposes was extracted from the course unit description of English Language I, Technical English I, Technical English II, and English for Management, which can be downloaded from [www.cud.upct.es](http://www.cud.upct.es). These course unit descriptions provide information about theoretical content, competencies and learning outcomes, learning goals, teaching methodology, and the assessment, where it is clearly stated that listening is a relevant aspect which has a value of over 32 percent in the final marks.

The initial skill of students in English as a foreign language is an input that may help the lecturers in charge to propose new content in order to give students specific guidance towards achieving a particular proficiency level.

The proficiency evolution trend was calculated by analysing the initial placement test that takes place at the beginning of the courses and the marks at the end and by putting the students into five groups following the Cambridge English Qualification scale.

This scale is as follows: "0" or Lv = 0.2 for students without a previous certification in English or with the lowest level of the receptive oral skill, "B1" or Lv = 0.4 for students with a minimum certificate or low English proficiency in the aforementioned skill, "B2" or

Lv = 0.6 for students with a middle certificate or middle proficiency level, “C1” or Lv = 0.8 for students with a high certificate or high proficiency in the skill, and “C2” for students with the highest certificate or highly proficient level on the abovementioned test. The probability distribution for each course is indexed in Table 1.

**Table 1.** English Level (t) Probability distribution for each Course.

English Level		Course1 (x = 1)	Course2 (x = 2)	Course3 (x = 3)	Course4 (x = 4)
0	(Lv = 0.2)	2 ( $p = 0.067$ )	1 ( $p = 0.033$ )	0 ( $p = 0.00$ )	0 ( $p = 0.00$ )
B1	(Lv = 0.4)	14 ( $p = 0.467$ )	8 ( $p = 0.267$ )	1 ( $p = 0.033$ )	0 ( $p = 0.00$ )
B2	(Lv = 0.6)	11 ( $p = 0.367$ )	17 ( $p = 0.567$ )	23 ( $p = 0.767$ )	14 ( $p = 0.467$ )
C1	(Lv = 0.8)	2 ( $p = 0.067$ )	3 ( $p = 0.100$ )	5 ( $p = 0.167$ )	10 ( $p = 0.333$ )
C2	(Lv = 1.0)	1 ( $p = 0.033$ )	1 ( $p = 0.033$ )	1 ( $p = 0.033$ )	6 ( $p = 0.200$ )

The proficiency evolution trend was calculated by applying quadratic regression to the probability distribution table, taking into account that students must cross a certain level to pass the courses, as illustrated by the fact that there must not be any student with Lv = 0.2 in the third course or students with Lv = 0.4 in the fourth course. Therefore, the number of students with a categorised level was calculated using five equations, one for each level, and an “x” variable that represented the course ( $x = 1, 2, 3, \text{ or } 4$ ; see Equations (1)–(5)), the coefficients of determination, “ $R^2$ ”, between 0.98 and 0.99 represent a good fit of the forecast value for the number of students with a specific English level in each course.

$$F(Lv = 0.2)_{LEVEL=0} = \frac{1}{30}(0.25x^2 - 1.95x + 3.75); 1 \leq x \leq 2 R^2 = 0.9818 \quad (1)$$

$$F(Lv = 0.4)_{LEVEL=B1} = \frac{1}{30}(1.25x^2 - 11.15x + 24.25); 1 \leq x \leq 3 R^2 = 0.981 \quad (2)$$

$$F(Lv = 0.6)_{LEVEL=B2} = \frac{1}{30}(-3.75x^2 + 20.25x - 6.25); 1 \leq x \leq 4 R^2 = 0.8571 \quad (3)$$

$$F(Lv = 0.8)_{LEVEL=C1} = \frac{1}{30}(x^2 - 2.4x + 3.5); 1 \leq x \leq 4 R^2 = 0.9947 \quad (4)$$

$$F(Lv = 1)_{LEVEL=C2} = \frac{1}{30}(1.25x^2 - 4.75x + 4.75); 1 \leq x \leq 4 R^2 = 0.9818 \quad (5)$$

This level distribution function allowed for a random database to be built with 30 students per course repeated 7 times, creating almost 210 records per course; in total, 820 records represented the 4 courses.

## 2.2. Methodology

The MWA-IT model [15] is based on the information theory, where workload is associated with a greater amount of processed information per unit of time, in addition to some factors that may increase or decrease its results, such as the task complexity or the subject’s experience or previous skill achievement in English listening practice.

With a focus on the listening task, the model was then applied to determine the rate of information processed (CV) as a mental workload indicator (see Figure 2). For such purposes, it was necessary to calculate the listening task complexity (ST), the physical variable measured as the median range of response time per question (MQ), and the experience variable, which was defined as the English level (LV) of each student.

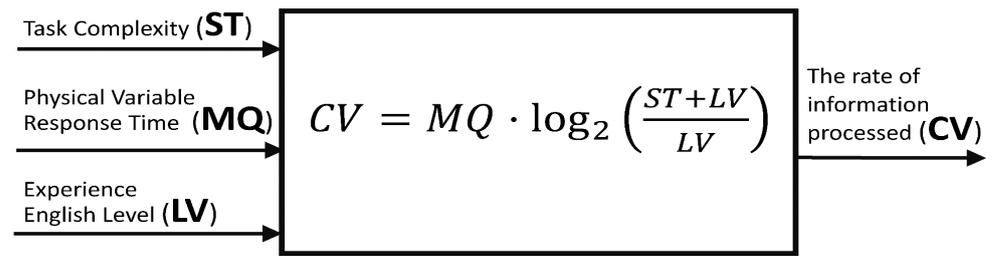


Figure 2. MWA-IT model. Source: adapted from [15].

With the MWA-IT model being used as a reference, the listening task complexity (*ST*) was defined as a variable that was dependent on the length of the conversation (*L*) expressed in seconds and within the limits of 120 to 300 s; the number of speakers involved (*S*), with a maximum limit of five; the English accent (*ACC*) used by the speakers (see Table 2); and the clarity of the conversation over a background of controlled noise (*CL*) (see Table 3). *ST* can be mathematically expressed as follows:

$$CV = MQ \cdot \log_2 \left( \frac{ST + LV}{LV} \right) \tag{6}$$

$$ST = \log_2 \left( \frac{L \cdot S \cdot ACC}{CL} \right) \tag{7}$$

where *ST* = Task Complexity; *MQ* = Physical Variable Response Time, *LV* = Experience English Level; *ACC* = Accent Listening Factor; *CL* = Clearly Listening Factor.

Table 2. Accent Listening Factor.

ACC	Description
0.7	Speakers use a Clear British Accent
1	At least one speaker uses a different accent e.g., Irish, north American, or any other similar accent
1.3	More than two speakers using accents or at least one strong accent such as Australian

Table 3. Clearly Listening Factor.

CL	Description
1	Noiseless environment (0% noise vs. 100% conversation volume)
0.8	Low and rhythmical Noise i.e., Natural Sound (10% noise vs. 90% conversation volume)
0.6	Soft Traffic environment, pedestrian street sounds (20% noise vs. 80% conversation volume)
0.4	Heavy traffic sounds (30% noise vs. 70% conversation volume)
0.2	High noise in bar or restaurant (40% noise vs. 60% conversation volume)

The working database to calculate the listening task complexity (*ST*) was compiled using a random combination of variables using the Excel function “randbetween(a,b)”, where “a” is the minimum value of the variable and “b” is the maximum. Therefore, it was possible to create a list of 40 different listening complexity levels plus 2 more for the minimum possible result (set up with a listening length of 120 s and 1 speaker with a clear British accent in a noiseless environment) and the maximum (set up with a listening length of 300 s and 5 different speakers with 1 strong accent or more than 2 speakers with an accent in a highly noisy environment). The results were arranged in ascending order to assign the lowest levels to the first courses (see Table 4).

**Table 4.** Listening task complexity (ST).

Course	Order	L	S	ACC	CL	ST	Course	Order	L	S	ACC	CL	ST
First Course	0	120	1	0.7	1	6.392	Third Course	21	256	2	1.3	0.6	10.115
	1	150	1	1.3	0.8	7.929		22	254	2	1.0	0.4	10.311
	2	127	1	1.3	0.6	8.104		23	266	2	1.0	0.4	10.377
	3	269	1	0.7	0.6	8.294		24	298	3	0.7	0.4	10.611
	4	246	1	0.7	0.4	8.750		25	167	2	1.0	0.2	10.706
	5	126	2	0.7	0.4	8.785		26	255	4	0.7	0.4	10.802
	6	195	2	1.0	0.8	8.929		27	289	5	1.0	0.8	10.819
	7	211	2	1.0	0.8	9.043		28	282	2	1.3	0.4	10.840
	8	247	1	1.3	0.6	9.064		29	149	5	1.0	0.4	10.863
	9	160	2	0.7	0.4	9.129		30	155	4	1.3	0.4	10.977
	10	183	1	1.3	0.4	9.216		31	295	2	0.7	0.2	11.012
Second Course	11	255	3	0.7	0.8	9.387	Fourth Course	32	213	3	1.3	0.4	11.020
	12	147	4	0.7	0.6	9.422		33	231	5	1.0	0.4	11.496
	13	282	2	1.0	0.8	9.461		34	248	3	1.0	0.2	11.861
	14	136	3	0.7	0.4	9.480		35	186	4	1.0	0.2	11.861
	15	120	3	1.3	0.6	9.607		36	274	4	0.7	0.2	11.905
	16	126	2	1.3	0.4	9.678		37	208	4	1.0	0.2	12.022
	17	169	2	1.0	0.4	9.723		38	287	5	1.3	0.4	12.187
	18	169	2	1.0	0.4	9.723		39	165	5	1.3	0.2	12.389
	19	249	2	0.7	0.4	9.767		40	194	5	1.3	0.2	12.622
	20	251	1	0.7	0.2	9.779		41	300	5	1.3	0.2	13.3

L = Listening Length (s); S = # of speakers; ACC = Accent (Table 2); CL = Clearly (Table 3).

The average range of the response time per question (MQ) when applied to a listening practice test was selected as the physical variable because of its relationship with the management of information extracted from the listening exam, where students who completely understood the listening comprehension required the lowest time to answer each question. This information was missing at the beginning of the study; however, this was overcome by using random data based on the assumption in Table 5, in agreement with the expectations of the teachers in charge, which filled the gap and enabled the completion of the analysis. The teachers indicated that a listening practice test should not last more than 2 h for a 30 question format. The time needed per question should be nearly 4 min for an average-level student, 2 min for a high-level student, and probably nearly 8 min for the student at the lowest level (see Table 5).

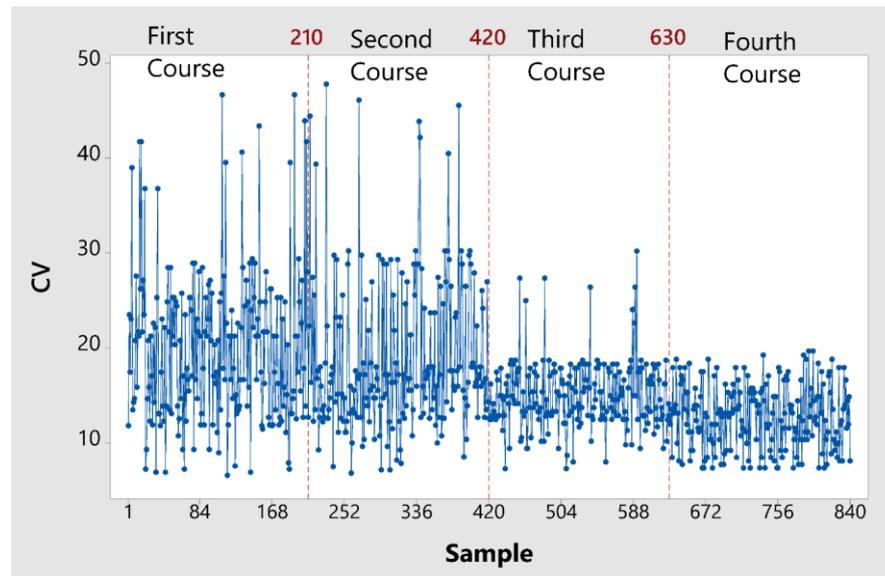
**Table 5.** Expected response time to listening questionnaires.

English Level	Lv	MQ
0	0.2	6.5–8.5 min
B1	0.4	4.5–6.5 min
B2	0.6	3.0–4.5 min
C1	0.8	2.5–3.5 min
C2	1	2.0–2.5 min

MQ = Median Range of response time per question. LV = English Level Factor.

The process of creating new data based on the experience of the teachers in charge allowed us to continue with the research, as the focus was to perform a workload study based on the MWA-IT and NASA-TLX methodologies and to set a new framework that may provide support for teachers’ decision-making processes. This limited the analysis of the results to a comparative framework rather than to an analysis of the absolute value of the results. However, the comparative outcome was still of interest, as it may suggest a new research line on EAP, where students’ workload indicators may play a relevant role in the teachers’ decision-making processes when designing new course unit descriptions.

Because of the assumptions regarding the  $MQ$  variable, it was possible to calculate the rate of information processed ( $CV$ ), depending on the specific complexity of each listening task. As there were ten different levels of complexity per course measured by the variable  $ST$  (see Table 4), ten different  $CV$  values for each sample (student) were calculated. The average of all of them was selected as the representative value for the rate of information processed by each student. The time series plot (see Figure 3) revealed lower  $CV$  values for the third and fourth courses, which can be explained by the fact that students acquire a higher level of English proficiency as they pass from one course to the next and that, in turn, this higher level of English must involve a lower cognitive workload.



**Figure 3.** Rate of information processed ( $CV$ ) plot from applying the MWA-IT model.

This result was analysed using MINITAB 19.0, which first revealed that this time series failed the normality test of Anderson–Darling, Ryan–Joiner, and Kolmogorov–Smirnov (see Figure 4). Afterwards, a multiple regression model was proposed to calculate the expected  $CV$  value (see Equation (8)) depending on the course and the level of English according to Table 1. The coefficient of determination  $R^2$  (0.9006) indicated that the regression model could finally be accepted.

$$\widehat{CV} = 58.104 + 1.759x_1 - 109.79x_2 - 0.2479x_1^2 + 57.99x_2^2 \quad (8)$$

where,  $x_1 = \text{Course}$ ;  $x_2 = \text{Level}$ ;

Returning to NASA-TLX, new variables had to be added to the working database in order to proceed with the study. NASA-TLX is calculated by applying an individual task load questionnaire after the task is performed. The questionnaire is focused on a rating scale for mental, physical, and temporal demand, together with effort, performance, and frustration levels, which must provide a single numerical result when finally evaluated.

Therefore, the formulated questions stand for the following: “mental” demand, to determine how much mental perceptual activity was required (thinking, deciding, searching, etc.); “physical” demand, which is more closely related to the physiological aspects caused by pushing, pulling, turning, controlling, etc.; “temporal” demand, to find out how much time pressure is felt due to the task; “effort” level, which focuses on how hard the worker has to work to accomplish the task; “performance” level, to determine how successful the worker feels after accomplishing the objectives; and “frustration” level, which defines how insecure, irritated, or stressed the workers are when the task is finished. Each rating scale is weighted to set up an index that represents the workers’ expectations regarding the complexity of the assigned task. It can ultimately be expressed as follows:

$$NASA_{TLX} = \frac{1}{n} \sum_{i=1}^n (M_i \cdot W_{Mi} + PH_i \cdot W_{PHi} + T_i \cdot W_{Ti} + PE_i \cdot W_{PEi} + EF_i \cdot W_{EFi} + FR_i \cdot W_{FRi}) \tag{9}$$

$W_{Mi}$  = weight representing the contribution of Mental Effort factor to the workload.

$W_{PHi}$  = “ the contribution of Physical demand factor to the workload.

$W_{Ti}$  = “ the contribution of Temporal demand factor to the workload.

$W_{PEi}$  = “ the contribution of Performance level to the workload.

$W_{EFi}$  = “ the contribution of Effort level to the workload.

$W_{FRi}$  = “ the contribution of Frustration Level to the workload.

$M_i$  = Averaged value of points marked in questions related to Mental Effort.

$PH_i$  = Averaged value of points marked in questions related to Physical demand

$T_i$  = Averaged value of points marked in questions related to Temporal demand

$PE_i$  = Averaged value of points marked in questions related to Performance level

$EF_i$  = Averaged value of points marked in questions related to Effort level.

$FR_i$  = Averaged value of points marked in questions related to Frustration level.

The questionnaire was completed with information collected from the surveys that the University Centre of Defence at the Spanish Air Force Academy compiled for each subject. It consisted of 12 questions (see Table 6) that could be linked to the NASA-TLX rating scale. However, as there were no previous publications or references concerning the expected weights of the mental, physical, temporal, performance, effort, and frustration level contributions to workload that applied to the “English Listening Learning” reference, the teacher in charge provided this information. The teacher stated that the mental factor was worth 40% of the total workload, the physical factors was worth 1%, the temporal and effort factors were worth almost 20% each, the performance factor was worth 15%, and the frustration factor was worth 4% of the total workload (see Table 7).

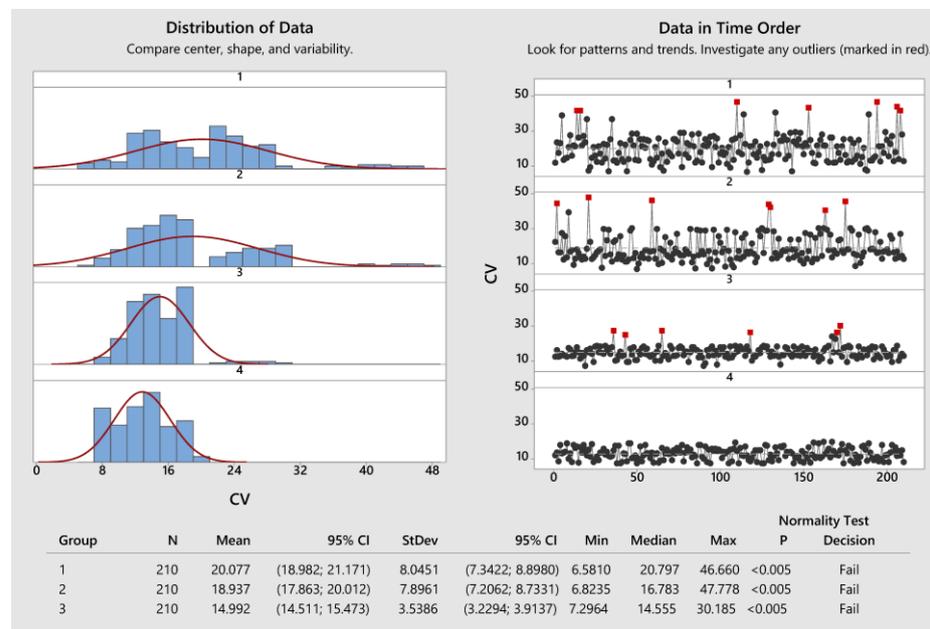


Figure 4. Minitab graphical summary of the CV and normality test.

**Table 6.** Task Load Questionnaire.

Description	
#1	The teacher increases the student interest in the subject.
#2	The teacher expositions are clear and fully understandable.
#3	The educational resources provided by the teacher are useful, what helps to pass the course.
#4	The teacher encourages the student to participate actively during the lessons.
#5	The teacher is fully accessible when the student needs to solve questions outside of the classes.
#6	The test results provided by the teacher help the student to know her/his progress evolution.
#7	The work assignments planned out of the classes are useful to the student.
#8	The teaching methodology helps the students to accomplish the learning process.
#9	The assessment methodology is appropriate.
#10	The student acquired the knowledge and abilities described in the course unit description.
#11	Generally, the student feels satisfied with the teacher's work.
#12	The student needs more time to answer each question.

**Table 7.** NASA-TLX Rating Scale Definitions.

Scale (Weights)	Description	Question #
M: Mental ( $W_M = 0.4$ )	How much mental and perceptual activity was required while listening to the conversation.	#04
Ph: Physical ( $W_{PH} = 0.01$ )	How much physical activity was required, i.e., if it was necessary to search for any other teaching material or extra lessons to acquire the expected knowledge.	#02 and #03
T: Temporal ( $W_T = 0.2$ )	How much time was needed to answer each question with regard to the listening test.	#12
PE: Performance ( $W_{PE} = 0.15$ )	How successful the student thinks that she/he was in accomplishing the goals of the listening test.	#06 and #09
EF: Effort ( $W_{EF} = 0.2$ )	How hard the student had to work to accomplish her/his level of performance.	#05, #07, and #08
Fr: Frustration ( $W_{FR} = 0.04$ )	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent the student felt during the listening and the test.	#01 and #10

This adaptation work performed for the questionnaires is a relevant limitation of the study with regard to the NASA-TLX assessment. However, the results provided are the best approximated values of the real NASA-TLX results that could be obtained at the time for this research. Therefore, the NASA-TLX results must be understood as an example of a methodology application rather than the case study characterisation.

The results were added to the working database in a column labelled "NASA-TLX". However, in keeping with the aim of the research, a new random function was used to calculate a new NASA-TLX value, labelled "NASA-TLX<sub>IND</sub>", based on a "what if" methodology where such weights could fluctuate between expected limits. For example, the mental weight was worth 30–50%, the physical factor 0–5%, the temporal demand 1–2%, and the performance level 15–30%; the remaining percentage was divided again, with 70% for effort and 30% for frustration level (see Equation (10)a,b).

Again, the results of NASA-TLX and NASA-TLX<sub>IND</sub> were first analysed using MINITAB 19.0 to perform the normality tests of Anderson–Darling, Ryan–Joiner, and Kolmogorov–Smirnov, which failed in all cases (see Figures 5 and 6). However, in this case, the developed multiple regression model had lower coefficients of determination:  $R^2 = 70.83\%$  for NASA-TLX (see Equation (11)) and  $R^2 = 65.63\%$  for NASA-TLX<sub>IND</sub> (see Equation (12)), these were accepted because the research assumption was based on the comparative framework, where errors in the absolute values were neutralised as they were expected to affect all the data equally.

$$W_{EF} = 0.7 \cdot (1 - (W_M + W_{PH} + W_T + W_{PE})) \tag{10a}$$

$$W_{FR} = 0.3 \cdot (1 - (W_M + W_{PH} + W_T + W_{PE})) \tag{10b}$$

$$\widehat{NASA}_{TLX} = 96.41 + 6.63X_1 - 166.7X_2 + 1.829X_1^2 + 85.33X_2^2 - 13.69X_1X_2 \tag{11}$$

$$\widehat{NASA}_{TLXIND} = 92.74 + 7.51X_1 - 163.7X_2 + 2.347X_1^2 + 87X_2^2 - 16.94X_1X_2 \tag{12}$$

$X_1$  = Course;  $X_2$  = Level;

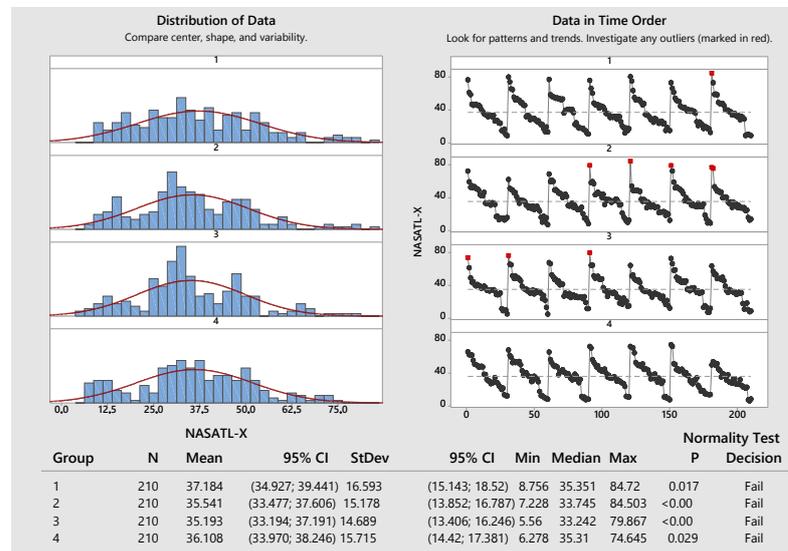


Figure 5. Minitab graphical summary of NASA-TLX and normality test.

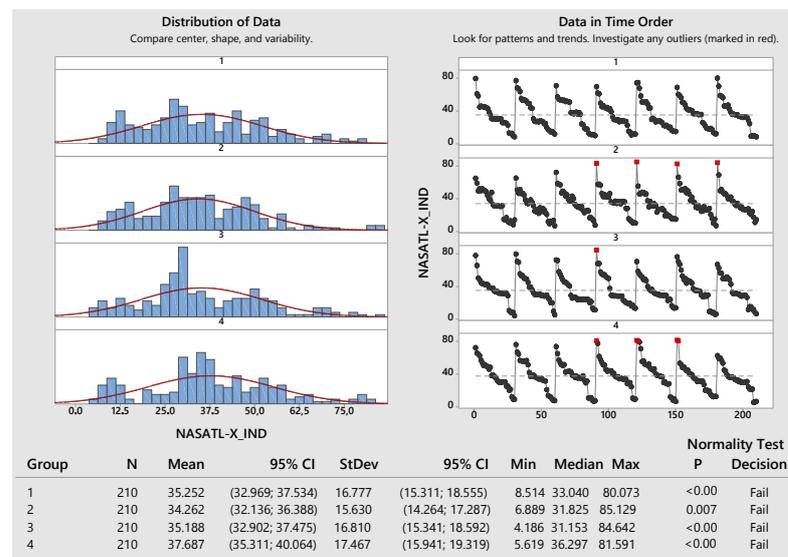


Figure 6. Minitab graphical summary of NASA-TLX<sub>IND</sub> and normality test.

### 3. Results and Discussions

The expected workload results derived from English listening training in combination with the students' initial level of English and their courses are summarised in Table 8, as is the expected progression towards the achievement of higher levels each year (levels from 0.2 to 0.8)

**Table 8.** Expected workload due English Listening Training depending on initial level of English.

		Initial Level = 0.2			Initial Level = 0.4				
Course	Level	$\widehat{NASA}_{TLX}$	$\widehat{NASA}_{TLXIND}$	CV	Course	Level	$\widehat{NASA}_{TLX}$	$\widehat{NASA}_{TLXIND}$	CV
1	0.2	72.2	69.9	40.0	1	0.4	46.4	44.3	25.0
2	0.4	53.0	52.0	26.0	2	0.6	31.3	29.9	15.6
3	0.6	38.8	39.0	16.2	3	0.8	21.2	20.5	10.4
4	0.8	29.6	30.8	10.5	4	1	16.1	15.9	9.4
	Total	193.7	191.7	92.6	Total		114.8	110.6	60.4
		Initial Level = 0.6			Initial Level = 0.8				
Course	Level	$\widehat{NASA}_{TLX}$	$\widehat{NASA}_{TLXIND}$	CV	Course	Level	$\widehat{NASA}_{TLX}$	$\widehat{NASA}_{TLXIND}$	CV
1	0.6	27.4	25.5	14.6	1	0.8	15.2	13.8	8.9
2	0.8	16.3	14.8	9.9	2	1	8.2	6.6	8.8
3	1	10.3	8.9	9.3	3	1	10.3	8.9	9.3
4	1	16.1	15.9	9.4	4	1	16.1	15.9	9.4
	Total	70.1	65.1	43.3			49.8	45.2	36.5

It can be seen in all cases that the CV index provided lower values than did those of NASA-TLX and NASA-TLX<sub>IND</sub>; in terms of a relative reference and as an example, the CV values were higher for the first and second courses. This circumstance is aligned with the expectations of the teachers in charge, as their aim is to increase the learning demand in such courses in order to decrease it later when the students start the third course with an intensive training programme. In a similar way, this is also aligned with the academic master plan expectations for the industrial engineering studies, where subjects with the highest scientific and technical demand are programmed to take place during the first two courses.

With the CV results, the expected workload is parametrised in terms of a CV rather than the interpretation of ECTs as hours of student dedication. Therefore, when a student starts with the lowest English level 1 (below B1 certification;  $L = 0.2$ ), the maximum workload caused by the four courses is expected to be 92.6; nearly 43% of such work occurs during the first course, 28% during the second course, 17% during the third course, and 12% at the end.

In the case in which the student might have started with a minimum certification in English (Level B1;  $L = 0.4$ ), the percentage distribution of the CV is similar, but the maximum workload in this case is 34% lower ( $CV = 60.4$  vs.  $CV = 92.6$ ). This reduction is more pronounced when the initial English level of the student corresponds to the certifications C1 ( $L = 0.6$ ) and C2 ( $L = 0.8$ ), where the reduction is nearly 50% ( $CV = 43.3$  vs.  $CV = 92.6$ ) and 60% ( $CV = 36.5$  vs.  $CV = 92.6$ ), respectively.

$$\widehat{NASA}_{TLX}R^2 = 70.83\%; \widehat{NASA}_{TLXIND}R^2 = 65.63\%; \widehat{CVR}^2 = 90.06\%.$$

Furthermore, in order to check whether the expected outputs correspond with the forecasted values and whether the validation data correspond to the observations, these results were used to perform hindcasting or back testing, which involves using known or closely estimated inputs into the developed models to test how well the outputs match the known results [41].

The hindcasting technique allows for several models to be compared by examining the mean square errors (Equation (13)) or the symmetric mean absolute percentage errors (Equation (14)) as a modification of these errors, where the divisor is half the sum of the actual and the forecasted values [42]. Thus, for each workload index

recorded, it was compared with the developed index by applying each forecasting model (Equations (8), (11) and (12)).

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \tag{13}$$

where  $\hat{Y}_i$  is a vector of n-predictions. and  $Y_i$  is the vector used as a reference.

$$SMAPE = \frac{1}{n} \sum_{t=1}^n \left( \frac{|F_t - A_t|}{\frac{F_t + A_t}{2}} \right) \tag{14}$$

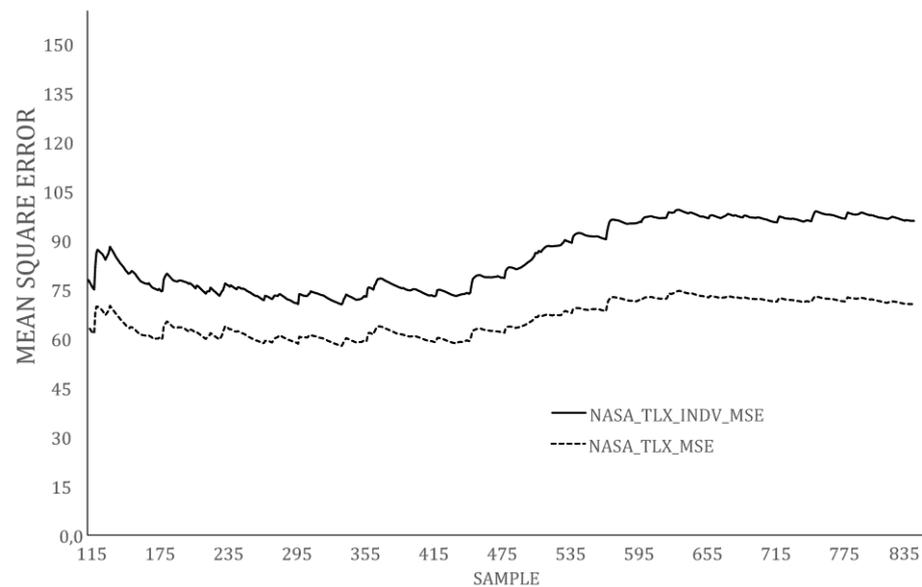
where  $F_t$  is the forecast value obtained with Vensim. and  $A_t$  is the value used as reference.

The MSE and SMAPE results indicated that the forecast model of the CV had the lowest error (see Table 9), while NASA-TLX and NASA-TLX<sub>IND</sub> had similar errors in relation to each other (see Figure 7a–c)). The rate of information processed could then be used as a valid cognitive workload; this might have been due to the assumptions mentioned throughout this research.

**Table 9.** MSE and SMAPE for the NASA TLX and CV workload indicators.

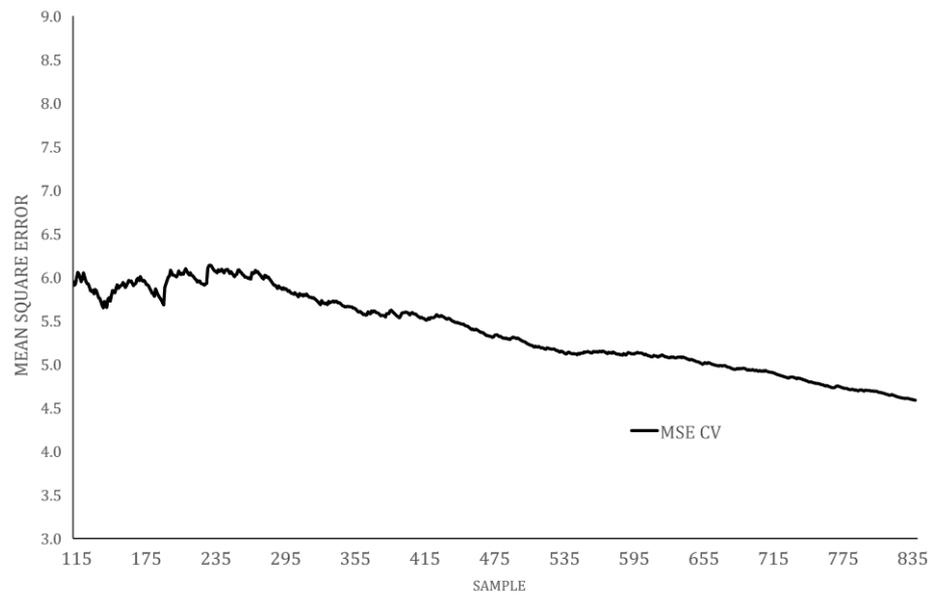
	MSE			SMAPE		
	NASA TLX	NASA TLX-IND	CV	NASA TLX	NASA TLX-IND	CV
<i>n</i> = 841 Av	70.443	95.86	4.586	0.202	0.238	0.113
Q3	72.122	96.845	5.741	0.1999	0.336	0.112
Q1	61.281	75.386	5.063	0.1751	0.088	0.108

Av = Average.

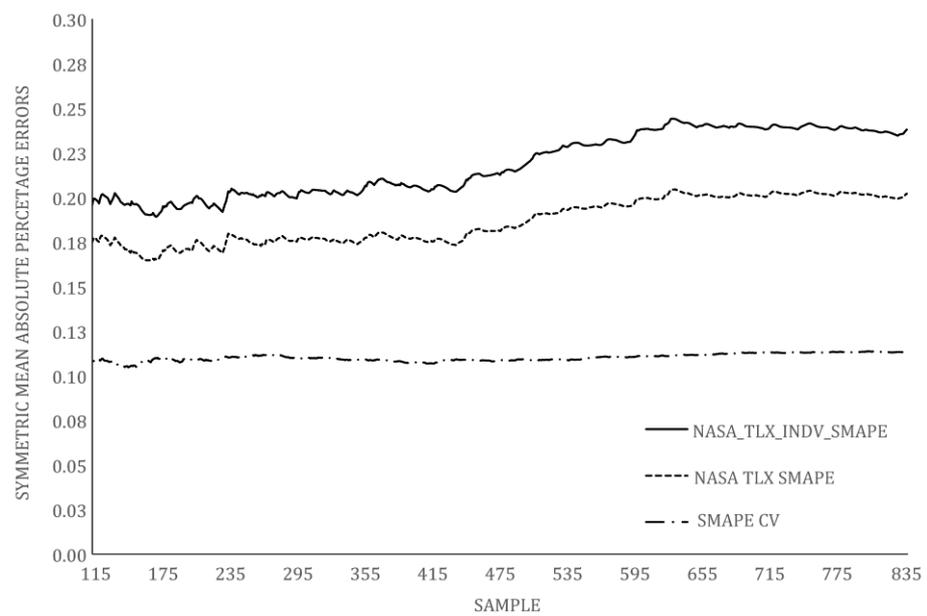


(a)

**Figure 7.** Cont.



(b)



(c)

**Figure 7.** (a) Mean square error of NASA-TLX and NASA-TLX IND. (b) Mean square error of CV per sample. (c) Symmetric mean absolute percentage or Workload indicators.

The MWA-IT model therefore allows the calculation of the lower and upper limits of the expected workload (CV: Q1–Q3); this calculation resulted in a nearly 40% less symmetric mean absolute percentage error than did the forecasting model for NASA-TLX.

In combination with the square root of the mean square error, it can be used to compare it with the maximum of the expected workload (See Table 8,  $L = 0.2$ ). The deviation can then be expressed in terms of the percentage of variance over the maximum workload. For example, for the upper limit of Q3 and the value of the square root of MSE ((CV:  $((5.741)^{0.5})/92.6 \cdot 100 = 2.58\%$ ; NASA-TLX:  $((72.122)^{0.5})/193.7 \cdot 100 = 4.38\%$ ) CV is a better model approach, as it again results in nearly 41% less deviation than does NASA-TLX.

Despite the analysis presented earlier, some research limitations must be recognised to complete the research scope statement. The first limitation is the questionnaire adaptation

from the case study to a NASA-TLX structure, as it was the only feasible way to add a subjective methodology into the research. In addition, the MWA-IT model was applied to a specific case study; however, it was not applied to a full subject but to a relevant part of English as a foreign language, the “Listening English Learning”. Nevertheless, the aim of the study was to apply this methodology to a learning process, which may encourage the development of other case study applications.

Finally, these results should be understood by students as a distribution percentage of the expected workload needed to pass the English listening program. Therefore, for initial levels of 0.2- 0.4, or even 0.6, it is expected that the students will need to use nearly 60-70% of their effort to pass the first two courses. As the military pilot training starts simultaneously with the third and fourth courses, the teachers in charge are therefore requested to increase the effort required at the beginning of the programme to avoid any additional disturbance later on.

#### 4. Conclusions and Future Works

The workload task index presents a new research perspective in the design of course unit descriptions; it creates a reference framework based on a numerical approach to assist the teachers in charge in making decisions about contents and difficulty levels when designing programmes in English for specific purposes (ESP) or English for academic purposes (EAP).

Regarding future workload studies in this field, it is of interest to use the MWA-IT model in combination with other workload methodologies for a comparative study when relevant data are missing. For such cases, this research presents a mathematical example of how missing data may be rebuilt to perform the study. However, there are still some variables that must be included for in-depth analysis, such as the complexity of the workload studies. Additionally, students’ marks could be used to identify specific workload values, not only to pass a course but to reach a specific level, which is a highly relevant matter to military students, who will earn future professional promotions according to their recorded academic results.

Rather than focusing on the limitations of the existing knowledge regarding this matter, it should be noted that the scope statement of the research was covered, providing a new framework with which to develop future case studies to delimit a subject workload. Additionally, the study also proves to be instrumental in determining the workload of a full degree by calculating and adding up the workload of the different subjects (in terms of ECTS) and thus taking into account the expected hours of autonomous study on the part of students, the hours of face-to-face classes, and the time devoted to the completion of assignments.

**Author Contributions:** The authors of this paper contributed equally to its revision before submission; the author contributions were as follows: J.L.R.-G., ideation, conceptualisation, methodology analysis, validation, and project management; J.-A.V.-L., hindcasting and statistical analysis, data curation, investigation, validation, and visualisation; M.N.P., resources, review on English as a foreign language for academic purposes, and data curation. The authors contributed equally to the other roles, such as writing, review, and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Data Availability Statement:** The research data will be made fully accessible upon request to the corresponding author.

**Acknowledgments:** The authors would like to thank the University Centre of Defence at the Spanish Air Force Academy and the Universidad de Castilla la Mancha for financial support. The authors would also like to express their sincere gratitude to the “Asociación Profesional de Ingenieros de Organización Industrial de España” due to the academic background information provided in the field of industrial engineering studies and STEM implications, as well as to the authors from the

Institute of Human Factors and Ergonomics at the College of Mechatronics and Control Engineering of Shenzhen and the school of Aeronautics of Northwestern Polytechnical University of Xi'an (China), whose work cited in this paper has inspired this research, opening new and interesting international collaboration opportunities regarding this matter.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Ke, X.; Yang, X. Research on English Testing for General Academic Purposes in General Engineering Universities. *Adv. Comput. Sci. Res.* **2017**, *1*, 736–739. [[CrossRef](#)]
- Zgłobiu, O.R. Filling The Gap Between Target Needs And Student Wants In English For Academic Environmental Purposes. *Stud. Univ. Babeş-Bolyai—Philol.* **2019**, *64*, 121–129. [[CrossRef](#)]
- Kuznetsov, A.; Krupchenko, A.; Schaveleva, E. Potential of an academic subject in profession-related competency formation: A case study of stakeholders' requirements in foreign language teaching within engineering education. In Proceedings of the INTED2017 Proceedings 11th International Technology, Education and Development Conference, Valencia, Spain, 6–8 March 2017; pp. 2801–2807. [[CrossRef](#)]
- Eksi, G.; Aydın, Y.C. English Instructors' Professional Development Need Areas and Predictors of Professional Development Needs. *Procedia—Soc. Behav. Sci.* **2013**, *70*, 675–685. [[CrossRef](#)]
- Peng, Z.E.; Wang, L.M. Listening Effort by Native and Nonnative Listeners Due to Noise, Reverberation, and Talker Foreign Accent During English Speech Perception. *J. Speech Lang. Hear. Res.* **2019**, *62*, 1068–1081. [[CrossRef](#)] [[PubMed](#)]
- Mansikka, H.P.; Virtanen, K.; Harris, D. Dissociation Between Mental Workload, Performance, and Task Awareness in Pilots of High Performance Aircraft. *IEEE Trans. Hum.-Mach. Syst.* **2019**, *49*, 1–9. [[CrossRef](#)]
- Ernawati, R.; Suhardi, B.; Pujiyanto, E. Using the NASA task load index and heart rate to evaluate vocational student's mental and physical workload. *AIP Conf. Proc.* **2019**, *2097*, 030057. [[CrossRef](#)]
- Patterson, J.T.; Hart, A.; Hansen, S.; Carter, M.J.; Ditor, D. Measuring Investment in Learning: Can Electrocardiogram Provide an Indication of Cognitive Effort During Learning? *Percept. Mot. Ski.* **2016**, *122*, 375–394. [[CrossRef](#)]
- Alshabeb, A.; Alsubaie, F.H.; Albasheer, A.Z. English for Specific Purposes: A Study Investigating the Mismatch between the "Cutting Edge" Book and the Needs of Prince Sultan Air Base Students. *Arab. World Engl. J.* **2017**, *8*, 376–391. [[CrossRef](#)]
- Huston, J.; Meier, S.; Faith, M.; Reynolds, A. Exploratory study of automated linguistic analysis for progress monitoring and outcome assessment. *Couns. Psychother. Res.* **2019**, *19*, 321–328. [[CrossRef](#)]
- Squires, A.; Miner, S.; Liang, E.; Lor, M.; Ma, C.; Witkoski Stimpfel, A. How language barriers influence provider workload for home health care professionals: A secondary analysis of interview data. *Int. J. Nurs. Stud.* **2019**, *99*, 103394. [[CrossRef](#)]
- Yang, H.-H.; Chang, Y.-H.; Chou, Y.-H. Subjective measures of communication errors between pilots and air traffic controllers. *J. Air Transp. Manag.* **2023**, *112*, 102461. [[CrossRef](#)]
- Anwar, K.; Wardhono, A. Students' Perception of Learning Experience and Achievement Motivation: Prototyping English for Academic Purposes (EAP). *Int. J. Instr.* **2019**, *12*, 271–288. [[CrossRef](#)]
- Cigdem, H.; Ozturk, M.; Topcu, A. Vocational college students' acceptance of web-based summative listening comprehension test in an EFL course. *Comput. Hum. Behav.* **2016**, *61*, 522–531. [[CrossRef](#)]
- Zhang, X.; Qu, X.; Xue, H.; Zhao, H.; Li, T.; Tao, D. Modeling pilot mental workload using information theory. *Aeronaut. J.* **2019**, *123*, 828–839. [[CrossRef](#)]
- Meshkati, N.; Hancock, P.A.; Rahimi, M.; Dawes, S.M. Techniques in mental workload assessment. In *Evaluation of Human Work: A practical Ergonomics Methodology*, 2nd ed.; Taylor & Francis: Philadelphia, PA, USA, 1995; pp. 749–782, ISBN 978-0-7484-0084-3.
- Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Adv. Psychol.* **1988**, *52*, 139–183. [[CrossRef](#)]
- Hart, S.G. Nasa-Task Load Index (NASA-TLX); 20 Years Later. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, San Francisco, CA, USA, 16–20 October 2006; pp. 904–908. [[CrossRef](#)]
- Pauzié, A. Evaluation of the driver's mental workload: A necessity in a perspective of in-vehicle system design for road safety improvement. *Cogn. Technol. Work* **2014**, *16*, 299–302. [[CrossRef](#)]
- Nikulin, C.; Lopez, G.; Piñonez, E.; Gonzalez, L.; Zapata, P. NASA-TLX for predictability and measurability of instructional design models: Case study in design methods. *Educ. Technol. Res. Dev.* **2019**, *67*, 467–493. [[CrossRef](#)]
- Azizah, N. The Packaging Section Employee Mental Workloads Analysis Using NASA-TLX Methods in Crunchy Peanuts Production. *Sci. Proc. Ser.* **2019**, *1*, 91–93. [[CrossRef](#)]
- Winter, J.C.F. de Controversy in human factors constructs and the explosive use of the NASA-TLX: A measurement perspective. *Cogn. Technol. Work* **2014**, *16*, 289–297. [[CrossRef](#)]
- Eichinger, A.; Bengler, K. Representations and operations: Parts of the problem and the solution. *Cogn. Technol. Work* **2014**, *16*, 307–310. [[CrossRef](#)]
- De Waard, D.; Lewis-Evans, B. Self-report scales alone cannot capture mental workload. *Cogn. Technol. Work* **2014**, *16*, 303–305. [[CrossRef](#)]

25. Rubio, S.; Díaz, E.; Martín, J.; Puente, J.M. Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Appl. Psychol.* **2004**, *53*, 61–86. [[CrossRef](#)]
26. Yang, X.; Kim, J.H. Measuring Workload in a Multitasking Environment Using Fractal Dimension of Pupil Dilation. *Int. J. Hum.–Comput. Interact.* **2018**, *35*, 1352–1361. [[CrossRef](#)]
27. Pütz, S.; Mertens, A.; Chuang, L.; Nitsch, V. Physiological Measures of Operators’ Mental State in Supervisory Process Control Tasks: A Scoping Review. *Ergonomics* **2023**, *0*, 1–54. [[CrossRef](#)]
28. Hollands, J.G.; Spivak, T.; Kramkowski, E.W. Cognitive Load and Situation Awareness for Soldiers: Effects of Message Presentation Rate and Sensory Modality. *Hum. Factors* **2019**, *61*, 763–773. [[CrossRef](#)] [[PubMed](#)]
29. Abe, T.; Dar, F.; Amnattrakul, P.; Aydin, A.; Raison, N.; Shinohara, N.; Khan, M.S.; Ahmed, K.; Dasgupta, P. The effect of repeated full immersion simulation training in ureterorenoscopy on mental workload of novice operators. *BMC Med. Educ.* **2019**, *19*, 318. [[CrossRef](#)] [[PubMed](#)]
30. Aljamal, Y.; Prabhakar, N.; Saleem, H.; Farley, D.R. Can the Perceived Difficulty of a Task Enhance Trainee Performance? *J. Surg. Educ.* **2019**, *76*, e193–e198. [[CrossRef](#)] [[PubMed](#)]
31. Brill, J.C.; Gibson, A.M.; Lawson, B.D.; Rupert, A.H. Do Workload and Sensory Modality Predict Pilots’ Localization Accuracy? In Proceedings of the 2019 IEEE Aerospace Conference, Big Sky, MT, USA, 2–9 March 2019; pp. 1–9. [[CrossRef](#)]
32. Mansikka, H.; Virtanen, K.; Harris, D. Comparison of NASA-TLX scale, modified Cooper–Harper scale and mean inter-beat interval as measures of pilot mental workload during simulated flight tasks. *Ergonomics* **2018**, *62*, 246–254. [[CrossRef](#)] [[PubMed](#)]
33. Golonka, E.M.; Bowles, A.R.; Frank, V.M.; Richardson, D.L.; Freyrik, S. Technologies for foreign language learning: A review of technology types and their effectiveness. *Comput. Assist. Lang. Learn.* **2014**, *27*, 70–105. [[CrossRef](#)]
34. Xue, Y. The Content-based Instruction in Campaign English Teaching. *DEStech Trans. Soc. Sci. Educ. Hum. Sci.* **2018**, *1*, 296–1–296–3. [[CrossRef](#)]
35. Hammami, S.; Saeed, F.; Mathkour, H.; Arafah, M.A. Continuous improvement of deaf student learning outcomes based on an adaptive learning system and an Academic Advisor Agent. *Comput. Hum. Behav.* **2019**, *92*, 536–546. [[CrossRef](#)]
36. Kim, J.; Kim, E.G.; Kweon, S.-O. Challenges in implementing English-medium instruction: Perspectives of Humanities and Social Sciences professors teaching engineering students. *Engl. Specif. Purp.* **2018**, *51*, 111–123. [[CrossRef](#)]
37. Bahrami, V.; Hosseini, M.; Atai, M.R. Exploring research-informed practice in English for academic purposes: A narrative study. *Engl. Specif. Purp.* **2019**, *54*, 152–165. [[CrossRef](#)]
38. Huang, Y.-T.; Guo, M. Facing disadvantages: The changing professional identities of college English teachers in a managerial context. *System* **2019**, *82*, 1–12. [[CrossRef](#)]
39. Ke, F.; Carafano, P. Collaborative science learning in an immersive flight simulation. *Comput. Educ.* **2016**, *103*, 114–123. [[CrossRef](#)]
40. Truschzinski, M.; Betella, A.; Brunnett, G.; Verschure, P.F.M.J. Emotional and cognitive influences in air traffic controller tasks: An investigation using a virtual environment? *Appl. Ergon.* **2018**, *69*, 1–9. [[CrossRef](#)]
41. Morley, S.K.; Brito, T.V.; Welling, D.T. Measures of Model Performance Based On the Log Accuracy Ratio. *Space Weather* **2018**, *16*, 69–88. [[CrossRef](#)]
42. Kim, S.; Kim, H. A new metric of absolute percentage error for intermittent demand forecasts. *Int. J. Forecast.* **2016**, *32*, 669–679. [[CrossRef](#)]

**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 MDPI and/or the editor(s). MDPI 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.