International Journal of Educational Innovations
ISSN 3078-5677

International Journal of Educational Innovations
Volume 2, Issue 1, 148-169
https://doi.org/10.46451/ijei.260112

Received: 24 September, 2025
Accepted: 20 October, 2025
Published: 1 January, 2026

Assessing Collocational Knowledge of ESP Learners: Methodological Considerations and Test Design Challenges

Višnja Pavičić Takač*
University of Osijek, Croatia
(*Corresponding author. Email: vpavicic@ffos.hr)

Mihaela Tabak
University of Zagreb, Croatia

Abstract
This paper addresses the challenge of assessing collocational competence in the context of English for Specific Purposes (ESP), where targeting specific learner needs, including those pertaining to their lexical knowledge, is particularly pronounced. The test of collocational knowledge used in this study (cf. Ding et al., 2024) integrates two main strands of L2 collocation research: corpus-based analyses of learner language (Nesselhauf, 2005) and test-based investigations of receptive and productive collocation knowledge (Revier, 2009). Given the specific focus on traffic and transport, a specialized corpus (10,542 words) was compiled from ESP textbooks and online sources for transport-related studies. TermoStat (Drouin, 2003) was used to extract domain-specific terms based on specificity and frequency relative to a general-language corpus. Collocations were identified via Sketch Engine, and distractors were drawn from the British National Corpus using mutual information scores. The final test included three tasks: (1) a COLLEX-based receptive task (Gyllstad, 2007), (2) a CONTRIX-inspired task (Revier, 2009) measuring full collocational knowledge, and (3) a translation task assessing productive recall (Bahns & Eldaw, 1993; Gitsaki, 1999). Scoring accounted for both lexical and grammatical accuracy. The assessment of test quality centred on a selection of psychometric parameters, namely reliability, discrimination power, and difficulty indices. The methodological complexity of testing collocational knowledge, which considers frequency, syntactic structure, and semantic association (Makinina, 2018; Szudarski & Conklin, 2014), is discussed. By combining corpus-based item selection with structured task design, this study offers a transparent model for assessing ESP collocational competence, addressing both benefits and pitfalls for test developers.

Keywords
Collocational competence, English for specific purposes (ESP), collocation testing

Introduction

Collocational knowledge—an essential component of deep lexical knowledge—is crucial for achieving lexical and language competence in a foreign language. It encompasses both receptive (recognizing forms and understanding meaning) and productive (producing correct forms and meanings) dimensions.

Given the importance of aligning language instruction with learners' professional and academic needs, assessing collocational knowledge in an English for Specific Purposes (ESP) context poses a unique challenge and thus requires a more tailored approach based on detailed needs analysis. Among other things, ESP collocation tests must be designed to focus on the linguistic rather than subject-matter knowledge and must measure field-specific, relevant collocations.

Research on collocation testing in the field of ESP has generally followed two distinct approaches. The first is corpus-based and primarily investigates the features of collocations in L2 learners' written output (e.g., Siyanova & Schmitt, 2008; Laufer & Waldman, 2011). The second approach centers on controlled assessments of collocational knowledge, the effectiveness of instructional interventions, and the cognitive processing of collocations (e.g., Gitsaki, 1996; Wray, 2008; Yamashita & Yiang, 2010; Wolter & Gyllstad, 2011).

Within the ESP domain, corpus-based studies represent the dominant strand in collocation research. The focus of these studies is mainly on examining lexical bundles and collocations occurring in written specialized texts, and on building specialized collocation lists (Tognini-Bonelli, 2001; Miščin, 2012; Miščin & Pavičić Takač, 2013; Salazar, 2014; He, 2023). Some studies, however, focus on collocation testing (Moder & Halleck, 2009; Durrant, 2014; Miščin et al., 2024). While the importance of collocational competence in ESP is well acknowledged, few standardized tests have been developed for the ESP context.

Since ESP targets domain-specific communicative competence, testing collocational knowledge in ESP differs from testing collocations in general English as a second language (ESL). ESP collocation tests must rely on specialized corpora extracted from authentic texts, which results in custom, small-scale tests, thus limiting generalizability. This requires a closer investigation of the existing collocation testing practices and the potential for developing a test tailored to the specific needs of ESP learners.

The aim of this study is to bridge that gap and to leverage corpus methodology into test design, for testing collocation recognition and production in an ESP context, and to explore which test format is optimal for assessing collocational knowledge.

We first briefly discuss the phenomenon of collocations in ESP and then review relevant research on testing ESP collocational knowledge. In subsequent sections, we describe the methodology and present and discuss the results of the present study.

Collocations in ESP

The vocabulary of ESP is often described as specialized, technical, or semi-technical. However, analyses of technical texts have shown that only a small proportion of the vocabulary is highly specialized or technical, while the majority consists of semi-technical and polysemous words. These words acquire their specialized and technical meaning only within specific contexts or when combined with other words to form collocations. Fraser (2009) refers to them as "crypto-technical" words and emphasizes that they constitute the largest portion of collocations in

specialized discourse. These are most commonly structures composed of noun + noun or adjective + noun combinations.

Specialized collocations, also referred to as collocations in terminology, are defined as multi-word expressions that contain at least one term functioning as the collocational base (Patiño, 2014; L'Homme, 2018), with the term referring to a concept specific to a particular field of expertise (L'Homme & Bertrand, 2000). They typically have several distinctive features, including subject specificity, formal characteristics (such as frequency of occurrence and co-occurrence, morphosyntactic features, a degree of (non)compositionality—i.e., fixedness between the collocational elements, and semantic cohesion). Terminologically relevant collocations primarily include lexical collocations, particularly noun + verb, noun + adjective, or noun + noun combinations. The nominal component is typically a technical term and represents the base of the collocation. In noun + noun collocations, the base is usually a domain-specific term, while the collocate is a nominalized verb or adjective.

The criterion of compositionality, however, is applied to only some collocations and is not used to distinguish compounds from collocations as in traditional linguistic theory. As a result, many terminological dictionaries list compounds as noun collocations. This approach was also adopted in the present study, where noun + noun structures were treated as collocations.

Another notable feature of ESP collocations is the so-called "collocational series phenomenon" (Heid, 1999), where a set of domain-specific terms that are conceptually related share the same collocate. Examples of such a series in English for traffic and transport would be: *reduce traffic congestion*, *reduce traffic jam*, and *reduce gridlock*. This phenomenon is particularly significant for teaching ESP collocations, as it facilitates learners' acquisition of specialized collocational patterns.

Testing Collocational Knowledge: Research Review

Most collocational test studies fall into one of the three main types: recognition-focused, recall-focused, or production-focused. Recognition-focused studies take into consideration multiple cognitive steps involved in the recognition process, that is, reading or listening, locating the collocation in the mental lexicon, and deciding about it (Siyanova & Schmitt, 2008; Henriksen, 2013). Recall acts as a bridge between recognition and production. It requires retrieving the form or meaning of a word when prompted (Gyllstad, 2009), whereas recognition involves selecting the correct word from the options. While recall and recognition are easier to control and test, assessing free productive use of collocations is more complex due to variability and unpredictability of responses. Consequently, even studies setting out to assess spontaneous production (Revier, 2009; Talakoob & Koosha, 2017) often actually measure controlled recall or production.

In their systematic review of collocation measurements used in empirical studies on L2 English, Ding et al. (2024) identified six instruments and seven item formats for assessing receptive and productive collocation knowledge. The instruments detected in this study include: the word association test (WAT), COLLEX, COLLMATCH, the PHRASE test, the frequency-based collocation test, and CONTRIX. The seven item formats identified are multiple-choice, judgement task, matching, fill-in-the-blank, translation, sentence generation, and error correction. Five of the six instruments assess receptive knowledge of collocations, and only the CONTRIX assesses productive knowledge in a semi-productive format.

Research on collocational knowledge assessment employs different methodologies, each with varying degrees of reliability, validity, and construct coverage. Studies focusing on collocation recognition primarily use multiple-choice tests or acceptability judgement tasks.

Gyllstad (2007) developed COLLEX, a multiple-choice test incorporating acceptable and non-acceptable high-frequency verb + noun collocations, and later introduced COLLMATCH (2009), a grid-format test requiring participants to judge the acceptability of word combinations in English. Both instruments demonstrated high internal consistency (Cronbach's $\alpha = 0.89$), indicating strong reliability. However, while COLLMATCH measures knowledge of whole collocations, COLLEX is limited to partial recognition. Similarly, Wolter and Gyllstad (2013) employed an acceptability judgement task for adjective + noun collocations across frequency levels, while Nizonkiza (2015) used a multiple-choice format for verb + noun collocations, achieving good reliability ($\alpha = 0.84$). These recognition-focused methods are generally effective for measuring receptive knowledge, yet they are susceptible to guessing and may not capture the depth of collocational knowledge.

Recall-oriented studies (Durrant & Schmitt, 2010) often employ fill-in-the-gap formats in which participants supply a node for an existing collocate, sometimes with minimal prompts (e.g., the first two letters). While these tasks test recall rather than recognition, their reliability and validity are rarely reported. Similarly, word association tests primarily assess associative links between words rather than comprehensive collocational knowledge, which limits their construct validity.

Testing productive knowledge typically involves controlled production tasks, often through translation from L1 to L2 or sentence-completion activities. For example, Kim (2017) used fill-in-the-blank translation tasks, whereas Revier (2009) designed an instrument requiring participants to match verbs with determiners and nouns in gap-fill sentences.

Revier's CONTRIX demonstrated very high reliability (Cronbach's $\alpha = .89$), confirming measurement precision. Validation studies suggest that it can distinguish collocational knowledge across groups. However, nonsignificant differences on some subtests indicate inconsistency in learners' ability to recognize semantic properties of collocations, and there is limited evidence of learners' awareness of the psychological reality of collocations.

Overall, while most instruments report strong reliability, their validity and construct coverage vary considerably. One of the few more recent studies comparing different task types for assessing L2 collocation knowledge was conducted by Lee and Shin (2021). To test 205 ESL learners' knowledge of collocations in academic English, the authors used four tasks (sentence writing, fill-in-the-blank, multiple-choice, and Yes/No acceptability judgement task). They found that task type had a strong effect on results, while collocation type had a minimal effect once frequency was controlled. The present study also compares various task types to assess collocation knowledge, but the design differs in terms of task types, morphosyntactic types of collocations included, and ensuing analysis. The decisions about the design were constrained by the features of the language of traffic, which is highly discipline-specific.

Methods

Test design

Selection of collocations

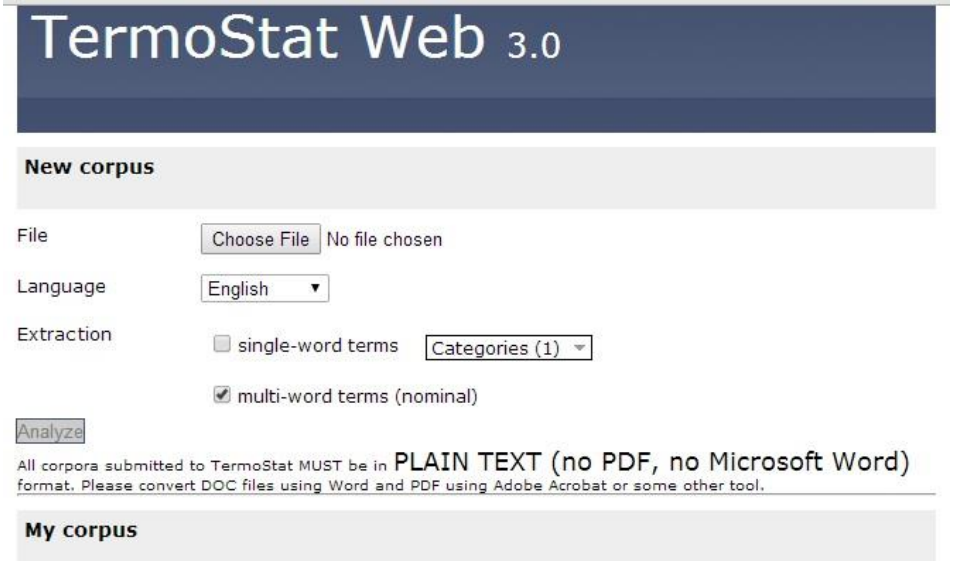
The selection of collocations for the test was initially based on the extraction of specialized terminology from a domain-specific corpus, followed by the identification of collocations

 ISSN 3078-5677

occurring with those terms. For the purposes of this research, a specialized corpus was compiled, consisting of selected texts from the English Textbook of Transport 1 (Bjelobrk & Bošković Gazdović, 2004), as well as several texts retrieved from various transport-related websites. These texts are used in the first-year curriculum for the courses English for Specific Purposes 1 and 2.

The corpus comprised a total of twenty texts (10,542 words; 12,095 tokens) introducing basic specialized vocabulary related to the following transport fields: road transport, rail transport, water transport, postal services, air transport, urban transport, telecommunications, ITS, and logistics. The corpus size was considered adequate given common practices in ESP research, where specialized corpora are often relatively small but carefully targeted. Methods such as TermoStat (Drouin, 2003) have been successfully applied to domain corpora of modest size to extract terms and collocations by contrasting them with general English corpora. Specialized texts often have a high density of domain-specific vocabulary, so even a smaller corpus can yield sufficient collocational information. The corpus was processed using the computational linguistic tool TermoStat (Drouin, 2003), a free online tool that extracts domain-specific terminology from corpora, relying on both statistical and linguistic analysis to identify potential terminological candidates (Figure 1).

Figure 1
Initial Interface of the Computational Linguistic Tool TermoStat (Drouin, 2003)



TermoStat analyzes the corpus by comparing the frequency of individual specialized terms within the input specialized corpus against a general language reference corpus embedded in the system. The reference corpus consists of 13,746 newspaper articles covering a broad range of topics and contains 7,400,000 tokens. By comparing the frequency of specialized terms with their frequency in the reference corpus, it is possible to observe the behavior of lexical units across the two corpora and identify lexical units that are characteristic of the specialized corpus.

The processing of the specialized corpus using the TermoStat computational linguistic tool yielded the following quantitative data: extracted lemmatized single-word and/or multi-word specialized terms, lemmatized terms ranked by frequency, statistically processed specialized terms based on their specificity, non-lemmatized token forms, terms categorized by part of speech, and contextual word analysis. Based on the specificity analysis, completed by TermoStat, the most frequent single- and multi-word units (i.e., those occurring minimally 10

times) within the corpus were identified (Table 1). The higher the specificity, the more likely it is that a term is domain-specific.

Table 1
Most Frequent Single- and Multi-Word Specialized Terms in the Specialized Corpus

Item	***f***
phone	25
driver	25
road	24
traffic	19
people	19
car	16
mobile phone	14
accident	14
light	14
effect	13
driving	12
system	12
use	11
percent	10
code	10

TermoStat offers the option of selecting among four different types of statistical analysis: term specificity, chi-square value (χ^2), log likelihood, and log odds ratio. For the purposes of this study, we identified the single- and multi-word tokens characteristic of the language used in the field of transport (Figure 2).

Figure 2
Extracted Specialized Terms Ranked by Specificity to the Domain

List of terms | Cloud | Stat | Structuration | Bigrams

Candidate (grouping variant)	Frequency	Score (Specificity)
mobile phone	14	93.16
cargo	25	63.62
chunnel	7	59.32
airplane	6	53.66
congestion	18	53.38
use of mobile phone	5	51.96
carpool	5	51.96
logistic	9	48.98
traffic congestion	7	48.39
biodiesel	4	45.18
traffic regulation	4	45.18
traffic	39	43.44
ethanol	6	42.76
alternative fuel	5	38.68
fuel option	3	37.26
parcel shipment	3	37.26
zip code	3	37.26
base station	3	37.26
break bulk	3	37.26

 ISSN 3078-5677

Given that statistical processing may result in the omission of certain specialized terms or the inclusion of general vocabulary items in the output list, subsequent manual verification by the researcher is necessary. This process included cross-checking candidate terms against specialized authoritative sources and their contextual use within the corpus to confirm domain-specific meaning. General vocabulary items and non-terminological collocations were filtered out, while specialized terms were kept, based on the specificity criteria. This step was taken to address limitations of statistical processing and ensure terminological accuracy. No subject specialists were consulted during this process. However, the corpus was based on an ESP textbook for students of traffic and transport, whose texts had been chosen with input from lecturers in specialized courses, ensuring expert validation of content relevance and accuracy.

The most frequent non-lemmatized token forms were plural nouns and gerunds. The fourth column in the table shown in Figure 3 displays such forms.

Figure 3
Extracted Non-Lemmatized Token Forms Ranked by Frequency

Results

List of terms | Cloud | Stat | Structuration | Bigrams

Candidate (grouping variant)	Frequency	Score (Specificity)	Variants
mobile phone	14	93.16	*mobile phone* *mobile phones*
cargo	25	63.62	*cargo* *cargoes*
chunnel	7	59.32	*chunnel*
airplane	6	53.66	*airplane* *airplanes*
congestion	18	53.38	*congestion*
use of mobile phone	5	51.96	*use of mobile phones*
carpool	5	51.96	*carpool*
logistic	9	48.98	*logistics*
traffic congestion	7	48.39	*traffic congestion*
biodiesel	4	45.18	*biodiesel*
traffic regulation	4	45.18	*traffic regulations*
traffic	39	43.44	*traffic*
ethanol	6	42.76	*ethanol*
alternative fuel	5	38.68	*alternative fuel* *alternative fuels*
fuel option	3	37.26	*fuel options*
parcel shipment	3	37.26	*parcel shipments*
zip code	3	37.26	*zip code* *zip codes*
base station	3	37.26	*base stations*
break bulk	3	37.26	*break bulk*

The collocations used in the tests were extracted using the corpus analysis tool SketchEngine and selected based on the mutual information (MI) scores of the collocates. Mutual information is a statistical measure that helps determine the strength and fixedness of collocations, their associative strength, and allows for distinguishing between strong and weak collocations (Miščin, 2012).

The criteria for including collocations in the tests (Table 2) were based on features that previous studies have identified as relevant for collocational knowledge, namely frequency, co-occurrence rate, and type of collocation (Durrant & Schmitt, 2009; Li & Schmitt, 2010; Wolter & Gyllstad, 2013; Boers et al., 2014; Szudarski & Conklin, 2014; Makinina, 2018). In applying the frequency criterion, we also considered non-lemmatized plural forms and/or different verb

forms (infinitives or other verb tenses), as these may have influenced the overall frequency; such forms were therefore grouped together.

Table 2
Items in the Collocational Competence Test according to Morphosyntactic Type, Frequency, and Associative Strength

Collocation MST	**HF/HAS**	**LF/LAS**	**LF/HAS**
verb + noun	*get off the bus; reduce traffic congestion*		*use public transport; obey traffic regulations*
noun + verb	*trains running*		*congestion occurs*
adjective + noun	*high speed trains; alternative route; pedestrian crossing*	*driving safety; suburban growth*	*available capacity; fatal car accidents*
noun + preposition + noun			*smoothness of acceleration*
adverb + adjective	*easily accessible*		
verb + adverb			*drive recklessly*
preposition + noun			*behind the (steering wheel)*
noun + noun	*escape route*		*parking meter; slip road*
verb + adjective			*keep (the streets) clear*
TOTAL number of MST	**8**	**2**	**11**
TOTAL number of test items	**21**		

MST = morphosyntactic type
HF - high frequency of occurrence
HAS - high associative strength between collocates (co-occurrence frequency calculated using the mutual information (MI) score)
LF - low frequency of occurrence
LAS low associative strength between collocates
NB: Although certain collocations include noun phrases, we retained SketchEngine's original annotation, which marks them uniformly as nouns (N).

It is important to note that specialized terms often appear as noun + noun structures. However, there is no consensus in the literature about whether these are compounds or collocations. Lewis (2000) includes compounds (n + n) in the same category as noun + noun combinations in his classification of collocation types. We assumed that the theoretical distinction between compounds and collocations would not impact the test, nor would it be considered relevant to

the participants. Therefore, in this study, noun + noun structures are treated as noun + noun collocations.

Given that the specialized corpus used in this study is relatively small, all collocations occurring more than once were considered high-frequency. Since no absolute thresholds for these categories exist in the literature, the criteria for high and low frequency and associative strength were determined based on their mutual relationship. For example, in the case of collocations classified as low-frequency/high associative strength (LF/HAS), the mutual information score was significantly higher than the frequency value (i.e. *fatal car accidents*), while high-frequency/high associative strength (HF/HAS) collocates (i.e. *reduce traffic congestion*) had values at least three times greater in both categories compared to low-frequency/low associative strength (LF/LAS) collocates (i.e. *driving safety*). Combinations classified as high-frequency/low associative strength (HF/LAS) were excluded from the test because they represent very weak collocations that can easily be replaced by free word combinations.

The number of collocates in the test was not evenly distributed by collocation type because of the uneven distribution of collocation types within the specialized corpus. However, the test included at least one of each morphosyntactic type of collocations. The decision to include various types of collocations was grounded in the fact that the specialized corpus was too small to generate a sufficient number of particular type(s) of collocation, as well as in recommendations found in previous research that the scope of collocations be broadened (cf., Nguyen & Webb, 2017). The most common morphosyntactic type of collocation in the test was the adjective + noun combination, reflecting its higher frequency in the corpus.

Test task types and scoring

The selection of tasks depends on what level of knowledge is to be assessed. Laufer and Goldstein's (2004) framework explains how different types of lexical knowledge vary in strength and aspects depending on what is being tested. It distinguishes between productive and receptive lexical knowledge, with productive knowledge relating to understanding the form, and receptive knowledge involving the recognition of meaning. In this context, productive knowledge is seen as "more advanced" than receptive knowledge. Another distinction concerns whether the test format requires recall or recognition. Recall involves generating a response, whereas recognition involves choosing from given options. Finally, there is a hierarchy of four levels or strengths of lexical knowledge: (1) productive recall (the most advanced), (2) receptive recall, (3) productive recognition, (4) receptive recognition. Following this classification and based on the assumption that it applies to collocational knowledge as one aspect of lexical competence, three types of tasks were selected for the collocational competence test in this study: one to assess receptive recognition of collocations, a second to involve both receptive and productive recall, and a third to measure productive recall. Inclusion of three different tasks aimed at capturing a wider range of knowledge.

Task 1 was based on the standardized multiple-choice receptive collocation knowledge test COLLEX (Gyllstad, 2007). The underlying assumption is that receptive knowledge involves the ability to recognize two or more words that frequently occur together as established linguistic combinations, as well as understanding the meanings of these word combinations. This task offers three different collocation options, one of which is correct, while the other two serve as distractors consisting of a correct collocation base combined with collocates that are typically not used with that base (i.e., pseudocollocations). The three-option-item format mirrors the original COLLEX test. Moreover, Rodriguez's (2005) meta-analysis showed that

three-option items are optimal, as additional distractors do not improve quality but increase the likelihood of implausible alternatives, thereby rendering some distractors non-functional. The collocations are presented in sentence contexts, with vocabulary chosen to ensure that participants understand the sentences and focus on the collocations. The distractors were chosen from the British National Corpus using the Sketch Engine, based on synonyms or low co-occurrence frequency in the corpus (Figure 3, also see Appendix). Unlike the original COLLEX format, this task included multiple types of collocations to reflect the variety of collocational types in the specific field of traffic. Scoring for this task was binary (1 indicating a correct response and 0 indicating an incorrect (or no) response).

Figure 4
Example of an Item in Task 1

1. Eleven departures per day were on offer with _____ every 45 minutes.
a) trains driving b) trains riding c) trains running

Task 2 (Figure 5, also see Appendix), modelled after the CONTRIX test (Revier, 2009), assesses both receptive and productive knowledge. It was chosen because it tests knowledge of the entire collocation without relying on the L1, requiring participants to select both the base and the collocate. The format features a sentence with a gap to be filled with a complete collocation (e.g., verb + (article) + noun). Next to the sentence, there is a table with possible collocation components, divided into three columns. Each column has three lexical items of the same category, such as three verbs, three prepositions, or three nouns. The participants are then asked to select one lexical item from each column to build the collocation that best fits the sentence's context.

Figure 5
Example of an Item in Task 2

1. The snow forced us to take [ ]	-	suitable	route
	an	additional	path
	a	alternative	road

Although this format involves selecting lexical items, the task does not test solely receptive knowledge, as participants must create meaning by combining appropriate units and simultaneously complete the collocate grammatically with an article or preposition. In the original CONTRIX version, all components in the table can combine to form three collocations, but only one fits the sentence context. In the version used for this test, only one collocation is correct and contextually appropriate, while other elements in the matrix serve as distractors.
Scoring for this task was the same as for Task 1. Because productive knowledge is required to "construct" the correct collocation, the following errors received a score of 0: incorrect base, incorrect collocate, both base and collocate incorrect, and incorrect grammatical collocation component (article or preposition).

Task 3 (Figure 6, also see Appendix) tested productive knowledge of collocations and involved translating seven sentences from participants' L1 into English. Participants were instructed to translate only the collocation, with the sentence providing additional context. This method for assessing productive collocational knowledge has also been used in earlier studies (Biskup,

1992; Bahns & Eldaw, 1993; Gitsaki, 1999). The same scoring scale (1 – correct, 0 – incorrect) used in the second task was employed.

Figure 6
Example of Item in Task 3

1) Mnogi europski gradovi izgubili su populaciju jer je došlo do rasta predgrađa.

Participants
The study involved 152 first-year students (81% male) enrolled at the Faculty of Transport and Traffic Sciences at a Croatian university. The average participants' age was 19 (M = 19.53; SD = 1.03). All participants had previously studied English; the majority (69%) had studied it as a compulsory school subject, while a smaller group (14%) had also attended private language classes. As a university entrance requirement, all participants had completed the B2 level of the state school-leaving examination in English.

Procedure
The test was administered during regular English classes as part of the formative assessment. Participants were informed that the test results would be used for the purposes of the present study and were asked to provide their consent. A total of 152 participants agreed to take part. They were also informed that participation was voluntary and that they could withdraw from the study at any time. During administration, participants were permitted to complete the tasks in any order and to review or revise their responses. They were instructed not to use any external aids, such as dictionaries. They were given a 45-minute time limit to complete the test.

Results
Table 3 displays the descriptive statistics for the three tasks. All 152 participants completed the test. Performance was highest on Task 1 (M = 0.79, SD = 0.15), moderate on Task 2 (M = 0.48, SD = 0.26), and lowest on Task 3 (M = 0.11, SD = 0.15). Minimum scores ranged from 0 to 3 correct answers, while maximum scores ranged from 5 to 7 correct answers, with Task 3 having the lowest observed maximum. Test reliability was calculated separately for each task because of the different task types. Since the tasks were scored binarily, the Kuder-Richardson reliability coefficient (KR) was used. All three test item scales had KR scores below .70, indicating low internal consistency.

Table 3
Descriptive Statistics for Three Tasks

Statistics	TASK 1	TASK 2	TASK 3
N	152	152	152
k	7	7	7
M	.79	.48	.11
S	.15	.26	.15
LOW	3	0	0
HIGH	7	7	5
RANGE	5	8	6
KR-20	.27	.60	.48

k = number of items
KR-20 = Kuder-Richardson-20 reliability

 ISSN 3078-5677

Next, we examined each item based on its facility, discrimination, and morphosyntactic structure to better understand the measurement quality of each task, specifically to see how effectively the selected collocations functioned as intended (Table 4).

Table 4
Item Facility and Discrimination Statistics

	ITEM	MST	IF_{total}	IF_{upper}	IF_{lower}	ID
TASK 1	1 trains running	N + V	0.63*	0.94	0.36	0.58*
	2 got off the bus	V + Prep + N	0.46*	0.72	0.16	0.56*
	3 reducing traffic congestion	V + N	0.94	1	0.86	0.14
	4 high speed trains	Adj + N	0.99	1	0.96	0.04
	5 available capacity	Adj + N	0.93	1	0.8	0.2
	6 driving safety	Adj + N	0.77	0.94	0.54	0.4*
	7 pedestrian crossing	Adj + N	0.85	0.96	0.68	0.28
TASK 2	1 an alternative route	Adj + N	0.66*	0.92	0.28	0.64*
	2 using public transport	V + N	0.64*	0.82	0.46	0.36
	3 smoothness of acceleration	N + Prep + N	0.41*	0.76	0.16	0.6*
	4 easily accessible	Adv + Adj	0.52*	0.92	0.2	0.72*
	5 congestion occurs	N + V	0.34*	0.72	0.06	0.66*
	6 were driving recklessly	V + Adv	0.37*	0.74	0.04	0.7*
	7 a slip road	N + N	0.41*	0.6	0.18	0.42*
TASK 3	1 suburban growth	Adj + N	0.07	0.2	0	0.2
	2 behind the wheel	Prep + N	0.29*	0.64	0	0.64*
	3 escape route	N + N	0.007	0.02	0	0.02
	4 fatal car accidents	Adj + N	0.1	0.26	0	0.26
	5 obey traffic regulations	V + N	0.17	0.42	0	0.42*
	6 parking meter	N + N	0.04	0.12	0	0.12
	7 keep (the streets) clear	V + adj	0.13	0.28	0	0.28

MST = morphosyntactic type
IF = item facility
ID = item discrimination

The item facility (IFtotal) indicates the percentage of test-takers who answered an item correctly, helping to identify easy, acceptable, and difficult items. Item discrimination (ID) refers to the extent to which each item effectively distinguishes between higher and lower-performing learners. Items with an IF between 0.26 and 0.75 are considered appropriately difficult, and items with ID values greater than .40 are regarded as acceptable (Brown & Hudson, 2002). These are marked with an asterisk in Table 4 for easier reference.

As shown, Task 1 was the least challenging task, as it included five items that were too easy (*driving safety, available capacity, pedestrian crossing, high speed trains,* and *reducing traffic congestion*). However, three items effectively distinguished between high and low scorers, two of which also had an appropriate difficulty level (*trains running* and *got off the bus*). An inspection of errors revealed that most errors occurred on the items containing the collocations *trains running* and *got off the bus*, the only two items that were both of appropriate difficulty

and effective discrimination power. Participants frequently opted for distractors that reflected literal translations from their L1, such as **trains driving* and **got out off the bus*.

All items in Task 2, which followed the CONTRIX format, had appropriate difficulty and discrimination indices, making the task the most suitable for the purpose. Errors were typically related to either the grammatical component of the collocation or to the collocate itself. When errors involved the collocate, participants often selected distractors based on synonymy or failed to eliminate options that were grammatically incorrect, such as choosing **driving reckless* instead of the correct *driving recklessly*.

Finally, Task 3 showed the highest level of difficulty, with only one item meeting the criterion for acceptable difficulty (*behind the wheel*). Most difficult items still effectively discriminated among high and low scorers (*behind the wheel, obey traffic regulations, suburban growth, fatal car accidents,* and *keep the streets clear*). Among the difficult items, two, however, did not discriminate well (*escape route* and *parking meter*). Common errors in participants' translations included omitting the collocation entirely by providing a paraphrase, circumlocution, or an approximation (e.g., *rules* for *regulations*, *growing of suburbs* for *suburban growth*, or *maintain the streets accessible for traffic* for *keeping the streets clear*) providing a partial translation of the collocation (e.g., *car accidents with death* for *fatal car accidents*), as well as literal translations from the participants' L1 (e.g., *respect the rules* for *obey the rules*). For example, the item *behind the wheel* mostly induced errors in the choice of preposition, indicating that participants were able to produce the conceptual meaning but struggled with grammatical accuracy.

Regarding the morphosyntactic type of collocations, it has been emphasized that certain types (e.g., adj + n and v + n) appeared more frequently in the test due to their natural occurrence in the corpus. Some morphosyntactic types (e.g., noun + preposition + noun) appeared only once. Therefore, the effect of the morphosyntactic type of the collocation cannot be assessed.

As for the effect of the frequency and association strength of the collocations (Tables 2 and 4), the results show that four of the eight high-frequency collocations with high association strength (i.e., HF/HAS) had acceptable levels of both difficulty and discrimination power. This set of collocations also included one difficult and two easy items. The low-frequency collocations with high associative strength (LF/HAS) also produced items with strong discrimination power (seven out of 11) and appropriate difficulty levels (six items). There were only two low-frequency items with low associative strength, of which one had acceptable difficulty and discrimination levels.

Discussion

The purpose of this study was to design a corpus-informed test that assesses learners' collocational knowledge through recognition and production tasks, and to determine which test format provides the most accurate measurement. All results must be interpreted with the specific study context in mind. The test had to be tailored to a highly domain-specific ESP field – that of traffic, which required compiling an adequate corpus. The resulting specialized corpus was relatively small, but it comprised collocations that participants would likely encounter in their professional use of English. Next, the test item selection had to meet the requirement of balancing collocations across the binary criteria of frequency, association strength, and morphosyntactic type, given the uneven distribution of collocation types within the specialized corpus. As some task types, such as multiple-choice, measure only receptive knowledge and do not consider the entire collocation, as one part of the collocation is always provided, it was

decided to include three task types to capture a broader array of collocation knowledge (from receptive to productive).

To evaluate the quality of the tasks, descriptive statistics were calculated. Kuder–Richardson (KR) reliability coefficient was low for all three tasks. The only task nearing borderline acceptability, with a value of .60, was Task 2. This was primarily due to the limited number of items in each task. Psychometric research indicates that reliability tends to decrease as the number of items decreases, as random error is enhanced when a construct is not sampled more extensively (Bachman, 1990). In our study, each of the three tasks contained only seven items, which may have constrained internal consistency values. The reliability of the tasks may also have been affected by the heterogeneity of the items, that is, the uneven distribution of collocation types across different morphosyntactic patterns, but also by the heterogeneity of the test takers (Bachman, 1990: 220), which is not unexpected in an ESP context at the tertiary education level. Therefore, low reliability coefficients should be interpreted cautiously, as they may reflect issues other than item quality or construct representation.

The mean scores suggest that Task 1 was the easiest (.79), and Task 3 the most difficult (.11). We now turn to a discussion of the factors that may account for these results.

Since the test was administered as part of a formative, norm-referenced learner assessment, item analysis included item facility and item discrimination (Bachman, 1990). Both analyses constitute a systematic statistical appraisal of test items, providing insight into the effectiveness of each item and into how appropriately it functions with a specific learner population within a given domain of interest (Brown & Hudson, 2002).

In Task 1, the multiple-choice recognition task assessing receptive knowledge, participants had to select the correct collocation from three options. The fact that five out of seven items in this task had the item facility index above .70 makes Task 1 overall easy. Gyllstad (2007) found that a three-option multiple-choice task was too easy for advanced learners; however, we speculated that other factors might be at play, given that the participants in our study were ESP learners and therefore not necessarily advanced in their specific domain. However, our attempt to discern potential causes of the observed facility from the features of collocations in this task, that is, their morphosyntactic type, frequency, association strength, transparency, congruency, and errors, failed, as no consistent patterns were revealed. We will revisit this issue after analyzing the remaining individual tasks. At this point, it appears that the Task 1 facility may be attributed to its format, specifically its requirement to recognize collocations.

The item discrimination analysis revealed that Task 1 contained three items with a good discrimination index. These items typically included distractors in the form of pseudocollocations or grammar-related alternatives (e.g., variation in suffix, preposition, or word class). Considering that two of the three items (i.e., *trains running* and *got off the bus*) were at the right difficulty level, the result suggests that these types of distractors may be more effective than synonym-based lexical alternatives. These findings are not entirely consistent with those of Lee and Shin (2021). In their study, the multiple-choice task proved to be the most effective at differentiating low and high scores. This was attributed to careful design and selection of as many as eight distractors, which reduced the likelihood of correct guesses. Additionally, the distractors mostly consisted of synonyms, near-synonyms, or plausible learner-generated word combinations among which participants had to identify the combination that was not only grammatically possible but also constituted a collocation in the target language. Based on the results of our study and those reported by Lee and Shin (2021),

it can be concluded that the multiple-choice format could be effective for assessing differences in participants' collocational knowledge if its design is based on careful selection of an adequate number of distractors. In a narrow ESP context, such as ours, generating eight plausible options may be challenging, if not impossible, but increasing the number of options seems worth further investigation.

Task 2 followed the CONTRIX grid format (Revier, 2009) and assessed receptive-productive collocation knowledge. It stands out with the highest number of items of appropriate difficulty (i.e., seven out of seven) and discrimination indices (six out of seven), suggesting that all its items function well. Most errors involved selecting an incorrect collocate or making mistakes in the grammatical elements of the collocation, such as articles, prepositions, or choosing the wrong word class (e.g., an adjective instead of an adverb). This indicates that learners generally recognized the target collocations but occasionally struggled to produce a complete and accurate collocation within the context. This raises the question of the role of grammatical knowledge in collocational competence. Poor performance may sometimes reflect weaknesses in grammatical knowledge rather than in lexical knowledge alone. Some scholars (Benson et al., 1986; Gitsaki, 1996; Sinclair et al., 2004) argue that errors in articles or prepositions should be considered collocational errors, while others (Hausmann, 1989; Nesselhauf, 2005; Wanner et al., 2006) maintain that collocations consist exclusively of lexical units and exclude grammatical combinations such as noun/verb + preposition. However, the distractor design, which included both semantic and grammatical alternatives, successfully highlighted these challenging aspects without making the items excessively difficult. This task format proved particularly effective for assessing both receptive and productive knowledge of collocations, as it required participants to form the complete collocation that fits the context. It seems safe to conclude that Task 2 was the most effective overall. Its strengths are its format (which included targeted collocations embedded in context, with attention to both meaning and form, i.e., lexical and grammatical components), integration of receptive and productive knowledge, and practical administration and scoring. It was also the only task that approached acceptable reliability levels, as indicated by the Kuder-Richardson coefficient, which aligns with previous studies (cf. Revier, 2009).

The use of a translation task is not uncommon in collocation studies (Laufer & Girsai, 2008; Peters, 2016; Zhang, 2017), especially if participants share the same L1. This was Task 3 in our study. It turned out to be the most difficult one, which was not unexpected, as it assessed productive collocation knowledge. Only one item had an appropriate level of difficulty, and only two items could clearly distinguish between stronger and weaker learners. The translation format required recall and production of target collocations in context. The analysis of responses showed that students often avoided translating the collocation, translated it only partially, or paraphrased it using familiar general words. Many answers also indicate a clear L1 influence. Since the collocations were incongruent between the two languages, learners could not rely on direct transfer, which likely made the task even more difficult. The Task 3 results suggest that translation tasks can capture productive collocation knowledge, but they also highlight the need for careful item design with attention to cross-linguistic factors in order to create tasks with a better balance of difficulty and discrimination.

As noted above, different morphosyntactic types of collocations were included in the test, among which some (e.g., adj + n and v + n) appeared more frequently, reflecting their specialized corpus frequency. This may have affected the reliability of the tasks, especially when a relatively small number of items is factored in for each task. Although the uneven spread of these types added heterogeneity across the three tasks, thus reducing reliability, the

selection of heterogeneous collocations made the test more valid because it reflected the variety of collocations learners meet in real ESP contexts. The results indicate that different morphosyntactic types of collocations contributed to variability in the test, but their effect on item quality was inconsistent and often depended on the task type. In our study, noun-based collocations (e.g., *escape route*, *parking meter*) were generally more predictable, whereas less typical structures such as verb + adverb (*drive recklessly*) or verb + adjective (*keep the streets clear*) showed greater variability in performance.

As for the influence of frequency and transparency of collocations on item facility and discrimination power, no clear patterns emerged. However, it was observed that in Task 1, high-frequency and transparent items, such as *get off the bus* or *high-speed trains*, were easier, while less frequent or less typical collocations, like *trains running* were harder. In Task 2, familiar structures such as *use public transport* were combined with less typical items like *smoothness of acceleration*, which increased task difficulty. Task 3 included easier noun-based collocations (e.g., *escape route*) as well as more demanding patterns such as *drive recklessly* or *keep the streets clear*. This variety of items within tasks targeting different knowledge levels helped to cover collocational knowledge more broadly.

A more comprehensive evaluation of collocational knowledge was also facilitated by including collocations with varying frequencies and association strengths. Given the small corpus, collocations that appeared more than once were considered high-frequency because they represent collocations that participants encounter in their ESP course. Thus, three groups of collocations were formed (Table 2): high-frequency and high association strength (HF/HAS), low-frequency and high association strength (LF/HAS), and low-frequency and low association strength. The HF/HAS collocations in the present test worked well overall, as many items had acceptable levels of difficulty and discrimination. Although a few items were too easy or too difficult, they contributed positively to test quality. The finding that high-frequency items are adequate was expected and aligns with Ellis et al. (2008), who noted a pronounced dependence on frequency in L2 learners. However, the best-performing items from this group were spread across the two tasks, namely the receptive and receptive–productive ones, which suggests that task type may have had a stronger effect than the collocation frequency. There were only two LF/LAS items in the test, each in a different task and with a distinct morphosyntactic structure. They call for revision as they were considered weak contributors in terms of both item facility and discrimination. The LF/HAS items performed surprisingly well, which is valuable because items with high discrimination but moderate difficulty tend to be quite informative. The favourable item discrimination in this set also shows that a strong association can compensate for low frequency. Nguyen and Webb (2017) emphasized the importance of node frequency in collocation recognition, but our data imply that association strength also plays a role, even when frequency is low. The implication that both frequency and association strength affect collocational knowledge lends support to previous studies (Durrant & Schmitt, 2009; Wolter & Yamashita, 2018), albeit with varying impact. This suggests that collocation knowledge is not only based on exposure to frequent collocations, but also on input that highlights specific collocations with high association scores, which is particularly significant in discipline-specific ESP contexts.

Taken together, and reiterating that these findings must be interpreted with caution, the above results provide evidence that features of collocations, such as frequency and association strength, affect test-takers' performance; however, they may be less influential than the task format itself. Therefore, our recommendation for an effective discipline-specific collocation test would be to select collocations based on an analysis of a specialized corpus to ensure their

relevance to participants, and to include both high- and low-frequency collocations with varied association strengths, as both attributes have a meaningful effect. Additionally, we recommend using the CONTRIX format, as it effectively measures receptive-productive collocational knowledge, is sensitive to differences in learners' collocational knowledge, and is practical to administer, provided the number of items is increased to enhance its reliability.

Conclusion

The present study demonstrated that designing a corpus-informed test of collocational knowledge in a highly domain-specific ESP field is a challenging yet valuable endeavour. The test format, which combined receptive, receptive–productive, and productive tasks, allowed for a multidimensional assessment of collocational knowledge. The analyses of item facility and discrimination facilitated the evaluation of the reliability and validity of the test by identifying items that are adequately challenging and effectively discriminate between different levels of performance.

Overall, the results suggest that the tasks varied in difficulty and discriminability, with Task 2 providing the greatest spread of performance and Task 3 potentially being too challenging for most participants. The data demonstrated that the most powerful tool for assessment of collocation use in specialized contexts was achieved by the task modelled on CONTRIX (Revier, 2009), which tapped into both receptive and productive knowledge. The findings also suggest that assessing collocational knowledge includes both lexical and grammatical components, emphasizing the importance of considering grammatical aspects when evaluating learners' command of collocations. However, reliability and validity analyses suggest that while the test captured relevant aspects of collocational knowledge, the relatively small number of items per task inevitably influenced internal consistency. This can be remedied by increasing the number of items in the task.

Although the limitations of the present study may seem numerous, their recognition not only strengthens the transparency of the analysis but also advances our understanding of ESP collocation knowledge assessment, as it highlights key methodological considerations for the development and evaluation of collocation tests.

Item and distractor selection may have been affected by relying on a relatively small reference corpus with an uneven distribution of collocation types across various morphosyntactic structures and an insufficient distinction between low- and high-frequency collocations. This limitation decreased the ability to fully control frequency, association strength, and structural balance, which could have impacted item difficulty and test reliability. Although corpus compilation may be constrained by specific characteristics of individual disciplines that define language for specific purposes, future research should address these issues by using larger, more balanced corpora to select items and establish a clear system for choosing distractors.

Another potential limitation is the dichotomous scoring, which renders all output containing an error, regardless of whether it is substantive or minor and grammatical in nature, incorrect. A more detailed, stratified scoring system could provide a better insight into what aspects of collocations participants struggle with, especially in productive tasks. A future study may explore the impact of different scoring systems.

References

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press.

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, 21(1), 101-114.

Benson, M., Benson, E. and Ilson, R. (1986). *Lexicographic description of English*. John Benjamins.

Biskup, D. (1992). L1 influence on learners' renderings of English collocations: A Polish/German empirical study. In P. Arnaud & H. Bẽjoint (Eds.), *Vocabulary and Applied Linguistics* (pp. 85-93). Macmillan.

Bjelobrk, B., and Bošković-Gazdović, K. (1998). *Transport and Communications*. Fakultet prometnih znanosti, Zagreb.

Boers, F., Demecheleer, M., Coxhead, A., and Webb, S. (2014). Gauging the effects of exercises on verb–noun collocations. *Language Teaching Research, 18*(1), 54–74.

Brown, J. D., and Hudson, T. (2002) *Criterion-referenced Language Testing.* Cambridge University Press.

Ding, C., Reynolds, B. L., Szabo, C. Z., & Boone, G. (2025). Assessing English language learners' collocation knowledge: a systematic review of receptive and productive measurements. *International Review of Applied Linguistics in Language Teaching*, *63*(2), 833-856.

Drouin, P. (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology, 9*(1), 99–115.

Durrant, P., and Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics.* 47, 157–77.

Durrant, P., and Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research, 26*(2), 163–188.

Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations: A meta-analysis. *International Journal of Corpus Linguistics*, *19*(4), 443-477.

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42(3), 375–396.

Fraser, S. (2009). Breaking down the divisions between general, academic and technical vocabulary: The establishment of a single, discipline-based word list for ESP learners. *Hiroshima Studies in Language and Language Education, 12*, 151–167.

Gitsaki, C. (1996). *The Development of ESL Collocational Knowledge*. *PhD thesis*. The University of Queensland).

Gitsaki, C. (1999). Second language lexical acquisition: A study of the development of collocational knowledge. San Francisco: International Scholars Publications.

Gyllstad, H. (2007). *Testing English collocations: Developing receptive tests for use with advanced Swedish learners*. Lund University.

Gyllstad, H. (2009). Designing and evaluating tests of receptive collocation knowledge: COLLEX and COLLMATCH. In *Researching collocations in another language* (153–170). Palgrave Macmillan.

Hausmann, F. J. (1989). *Le dictionnaire de collocations in FJ Hausmann, HE Wiegand, and L. Zgusta (Ed): Wörterbücher, dictionaries, dictionnaires*. Ein internationales Handbuch zur Lexikographie.

He, M., & Ang, L. H. (2023). Profiling a microeconomics noun collocation list: A corpus-based approach. *Southern African Linguistics and Applied Language Studies*, *41*(2), 191-209.

Heid, U. (1999). Extracting terminologically relevant collocations from German technical texts. In *TKE* (Vol. 99, 241–255).

Henriksen, B. (2013). Research on L2 learners' collocational competence and development–a progress report. *C. Bardel, C. Lindqvist, and B. Laufer (Ed.) L, 2*, 29–56.

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, *29*(4), 694-716.

Kim, S. (2017). "Minimum input, maximum output, indeed!" Teaching collocations through collocation dictionary skills development. *Lexikos*, 27, 265–286

L'Homme, M. C., and Bertrand, C. (2000). Specialized lexical combinations: should they be described as collocations or in terms of selectional restrictions. In *Proceedings. Ninth EURALEX International Congress* (497–506).

L'Homme, M. C. (2018). Maintaining the balance between knowledge and the lexicon in terminology: a methodology based on Frame Semantics. *Lexicography*, *4*(1), 3–21.

Laufer, B., and Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399–436.

Laufer, B., and Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*(2), 647–672.

Lee, S., & Shin, S. Y. (2021). Towards improved assessment of L2 collocation knowledge. *Language Assessment Quarterly*, *18*(4), 419-445.

Lewis, M. (Ed.) (2000). *Teaching collocation. Further developments in the Lexical Approach.* Language Teaching Publications.

Li, J., and Schmitt, N. (2010). The development of collocation use in academic texts by advanced L2 learners: A multiple case study approach. *Perspectives on formulaic language: Acquisition and communication*, 22–46.

Makinina, O. (2018). Factors impacting collocation recognition and controlled production by ESL speakers. *PhD thesis*. Carleton University.

Miščin, E. (2012). Glagolske kolokacije u engleskom jeziku medicinske struke. *PhD thesis.*

Miščin, E., and Pavičić Takač, V. (2013). Analiza kolokacijske kompetencije neizvornih korisnika engleskog jezika medicine. *JAHR*, 4 (7), 235–256.

Miščin, E., Patekar, J., & Božinović, N. (2024). Corpus of collocations in business English and students'collocational competence. *Rithink*, *14*, 31-43.

Moder, C. L., & Halleck, G. B. (2009). Planes, politics and oral proficiency: Testing international air traffic controllers. *Australian Review of Applied Linguistics*, *32*(3), 25-1.

Nesselhauf, N. (2005). *Collocations in a learner corpus* (Vol. 14). John Benjamins Publishing.

Nizonkiza, D. (2015). Measuring receptive collocational competence across proficiency levels. *Stellenbosch Papers in Linguistics*, *44*, 125–146.

Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. Language Teaching Research, 21(3), 298–320.

Pati ño, P. (2014). Towards a definition of specialized collocation. *LSP in Colombia: advances and challenges, Linguistic Insights. Studies in Language and Communication*, 119–133.

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, *20*(1), 113–138.

Revier, R. L. (2009). Evaluating a new test of whole English collocations. In *Researching collocations in another language* (p. 125-138). Palgrave Macmillan.

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, *24*(2), 3–13.

Salazar, D. (2014). 1. Theory and applications of phraseology and lexical bundles. In *Lexical Bundles in Native and Non-native Scientific Writing* (pp. 9-40). John Benjamins Publishing Company.

Sinclair, J. M. (Ed.). (2004). *How to use corpora in language teaching* (Vol. 12). John Benjamins Publishing.
Salazar, E. (2014). Bayesian reference analysis for exponential power regression models. *Journal of Statistical Distributions and Applications*, *1*(1), 12.
Siyanova, A., and Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, 64(3), 429–458.
Szudarski, P., and Conklin, K. (2014). Short-and long-term effects of rote rehearsal on ESL learners' processing of L2 collocations. *TESOL Quarterly*, *48*(4), 833–842.
Talakoob, F., & Koosha, M. (2017). Productive and receptive collocational knowledge of Iranian EFL learners at different proficiency levels. *International Journal of Applied Linguistics and English Literature*, 6(7), 11–16.
Tognini-Bonelli, E. (2001). Corpus linguistics at work.
Yamashita, J., and Jiang, N. A. N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *Tesol Quarterly*, *44*(4), 647–668.
Wanner, L., Bohnet, B., and Giereth, M. (2006). Making sense of collocations. *Computer Speech & Language*, *20*(4), 609–624.
Wolter, B., and Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, *32*(4), 430–449.
Wolter, B., and Gyllstad, H. (2013). Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition*, *35*(3), 451–482.
Wray, A. (2008). *Formulaic language and the lexicon*. Cambridge University Press.
Zhang, X. (2017). Effects of receptive-productive integration tasks and prior knowledge of component words on L2 collocation development. System, 66, 156–167.

APPENDIX: TEST OF COLLOCATIONS IN THE FIELD OF TRAFFIC

TASK 1 Choose the correct answer for each question (ONLY ONE ANSWER PER QUESTION):

1. Eleven departures per day were on offer with _____ every 45 minutes.
a) trains driving b) trains riding c) trains running
2. They _____ at Holborn and got a train to Mile End.
a) got off the bus b) got out of the bus c) took off the bus
3. This park-and-ride facility now provides an excellent, free car park and will very usefully contribute to _____.
a) diminishing traffic congestion b) reducing traffic congestion c) lessening traffic congestion
4. Where _____ are used, average speeds may be over 160 km per hour.
a) great speed trains b) strong speed trains c) high speed trains
5. Traffic ultimately expands to fill _____ on the roads.
a) available ability b) available capacity c) available capability
6. Techniques to improve _____ or to reduce urban congestion have earlier focused on the development of new roads.
a) driving safety b) driving security c) driving reliability
7. Drivers and riders should always stop still at _____ .
a) pedestrian cross b) pedestrian crossing c) pedestrian pass

TASK 2: Combine the words in the boxes to create the correct phrase (ONE WORD FROM EACH COLUMN ONLY):

1. The snow forced us to tak	-	suitable	route
	an	additional	path
	a	alternative	road

2. More and more people are taking to their bikes as an alternative to driving or	riding	private	transport
	using	passenger	service
	taking	public	car

3. In the case of the trolley car for instance, the depends on the skill of the motorman as the control equipment is manually operated.	tenderness	of	acceleration
	softness	to	velocity
	smoothness	in	speed

4. The M4 motorway and A34 trunk road make Didcot by car.	an	easy	accessible
	as	easily	available
	-	likely	acceptable

5. The main reasons why are more cars, poor road management	a	congestion	appears
	-	overcrowding	begins
	the	clogging	occurs

6. Don't pretend it was all the other driver's fault, even if they did crash into the back of you	-	driving	iresponsible
	was	riding	reckless
	were	going	recklessly

7. They entered the dual carriageway from	-	lane	ramp
	the	slip	roadway
	a	drive	road

TASK 3: Translate only <u>the underlined part</u> of the sentence:
Ex. Svi sudionici u prometu moraju poštovati <u>prometne znakove.</u>
traffic sign

1) Mnogi europski gradovi izgubili su populaciju jer je došlo do <u>rasta predgrađa.</u>

2) Pobrinite se da sjedite udobno <u>za volanom</u>.

3) Zbog sigurnosnih razloga svaka zgrada mora imati <u>put za evakuaciju</u>.

4) U 2013. dogodilo se 32 719 <u>automobilskih nesreća sa smrtnim ishodom.</u>

5) Svi sudionici u prometu moraju <u>poštovati prometna pravila</u>.

 ISSN 3078-5677

6)Automat za naplatu parkiranja je izvrstan izbor za one koji trebaju parkirati kratko vrijeme.

7) Zadatak komunalne službe je održavati ulice prohodnima za prometovanje.

 ISSN 3078-5677