

Article

The Impact of Topic Selection on Lexico-Grammatical Errors and Scores in English Oral Proficiency Interviews of Korean College Students

Yongkook Won ¹  and Sunhee Kim ^{2,*} 

¹ Center for Educational Research, Seoul National University, Seoul 08826, Republic of Korea; purgatorio@snu.ac.kr

² Department of French Language Education, Seoul National University, Seoul 08826, Republic of Korea

* Correspondence: sunhkim@snu.ac.kr

Abstract: The primary objective of this study is to identify the types of errors made by Korean college students in an oral proficiency interview in relation to specific task topics, and to examine how these errors affect their lexico-grammatical proficiency scores. Ninety-six two-minute-long audio clips of 32 Korean college students on three different topics were transcribed. Lexico-grammatical errors were then coded for statistical analysis and lexico-grammatical scores were estimated using many-facet Rasch measurement analysis with two raters. Friedman tests and Wilcoxon signed-rank tests showed that noun phrase, verb phrase, and prepositional phrase errors were more frequently found with the descriptive tasks than compare-and-contrast or hypothetical prompts. A hierarchical multiple regression analysis showed that noun phrase and verb phrase errors accounted for 22% of the variance in lexico-grammatical scores. Adding utterance length variables to the initial regression model explained an additional 43% of the variance in the lexico-grammatical scores. These findings suggest that noun phrase errors and verb phrase errors should be a priority in English classes, and that it is beneficial to teach English speaking skills in a way that takes into account the task characteristics and contextual factors.

check for
updates

Citation: Won, Y.; Kim, S. The Impact of Topic Selection on Lexico-Grammatical Errors and Scores in English Oral Proficiency Interviews of Korean College Students. *Educ. Sci.* **2023**, *13*, 695. <https://doi.org/10.3390/educsci13070695>

Academic Editor: Lawrence Jun Zhang

Received: 30 May 2023

Revised: 30 June 2023

Accepted: 7 July 2023

Published: 9 July 2023



Copyright: © 2023 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/>).

Keywords: error analysis; EFL oral proficiency; college student; lexico-grammar; topic selection

1. Introduction

Errors that occur during the process of learning a foreign language serve as a measure of the learner's acquisition and development of the target language. Corder [1] claimed that the errors made by learners can be a learning strategy for acquiring a new language, rather than simply a negative signal. Through error analysis, learners can identify the areas where acquisition is not going well and find the direction for learning [2]. In addition, error analysis, which explains the accuracy levels, is used to predict test takers' proficiency levels. Accuracy, along with fluency and complexity, has been used as the primary predictor of language proficiency [3,4]. With the increasing importance of automated language assessment, these linguistic features are likely to be included in automated speaking assessments, and they can be used to provide adaptive feedback to test takers or language learners [5].

Numerous studies have emphasized the use of error analysis as a means of improving students' foreign language learning, with a particular focus on identifying the challenging aspects that students may encounter [6–8]. While most of these studies have concentrated on errors made in writing, only a small amount of research has been conducted on error analysis of the speaking of English language learners [9]. Furthermore, these studies rarely considered the topics of the input used.

In performance-based oral communication tests, prompts or inputs affect examinees' performances, leading to variation in proficiency scores [10,11]. While these studies focused

on the effect of tasks on test takers' performance in general, they did not examine specific language variations. For example, Won [12] investigated the effects of task complexity on test takers' linguistic outputs, such as the subordination index [13], Guiraud index [14], number of errors per AS-unit [15], speech rate, repairs, filled pauses, and their related proficiency scores. However, little attention has been paid to the effect of test inputs on the types of errors that test takers make in their responses in speaking tests.

The present study aims to identify the types of errors made by Korean college students in spoken English, depending on the task topics. Specifically, the study investigates the types of lexico-grammatical errors produced by college students, how these errors relate to the interview questions' topics, and how they affect students' proficiency levels. The research questions for this study are

- (1) What are the frequencies and types of lexico-grammatical errors made by Korean college students in an English oral proficiency interview, depending on the task topics?
- (2) How do lexico-grammatical errors affect the lexico-grammatical scores of Korean college students in an English oral proficiency interview, depending on the task topics?

2. Literature Review

This section begins with an introduction to the concept of error analysis (EA) and the categorization of error types for in-depth analysis. It then delves into the previous studies that employed EA to investigate the distribution of error types made by Korean students in spoken English. This section concludes by outlining the statistical methods employed in the analysis of the associated data.

2.1. Error Analysis

EA is a process that involves carefully analyzing the errors made by language learners to understand the process of second-language learning and to suggest suitable teaching and learning strategies and remedial measures necessary for the target language. The reason is that errors made by language learners have been viewed as a natural part of the language acquisition process, rather than a sign of a language filled with grammatical mistakes. Errors are inadequacies that originate from a lack of linguistic knowledge and are systematic, while mistakes are seen as performance-related errors [1]. Errors cannot be self-corrected until the learner receives more pertinent input, and errors, not mistakes, are important elements that have contributed to language acquisition research [16,17]. EA differs from contrastive analysis (CA) in that EA focuses on examining the mistakes made by students and identifying areas where improvement is needed while CA involves analyzing the differences between two languages to identify areas where learners might encounter difficulties [18].

Observing learner errors can thus help identify the language system being used and this is crucial for several reasons. First, error analysis allows teachers to evaluate how far a learner has progressed and what needs to be taught as they move forward. Second, learner errors provide researchers with evidence of how a language is learned and the strategies used by learners. Third, errors are a valuable tool for learners themselves in achieving target language proficiency [18]. However, error analysis also has several weaknesses. One significant problem is the difficulty of reliably classifying errors based purely on linguistic information. Moreover, error analysis is limited to analyzing the output of the learner, and cannot adequately consider receptive skills. The avoidance strategy, when a student decides not to use a form they are not confident in, is also not taken into account in error analysis [19,20].

2.2. Surface Strategy Taxonomy

Dulay, Burt, and Krashen [17] proposed a surface strategy taxonomy that categorizes errors into four categories: omission, addition, misformation, and misordering. These categories are based on the possibility of changing the surface structure of a morpheme or verb. Omission errors occur when a morpheme or word is missing from a well-formed

utterance [17]. This omission can occur with both content morphemes (e.g., nouns, verbs, adjectives, adverbs) and grammatical morphemes (e.g., function words), which happens more frequently [17,21]. For example, the sentence “I (am) wondering if it is really possible things” lacks the *BE-verb* after *I*, and the sentence “there is no idea to say (about) it” lacks the function word *about*. Addition errors occur when there is the presence of any morpheme or word that should not appear in a well-formed utterance. For example, the preposition *for* is wrongly inserted in the sentence “I read essay about for my future”. Misformation errors occur when an incorrect form of the morpheme or structure is used. For example, the expression “taking a house” is incorrectly used instead of “buying a house” in the sentence “we can save some money for taking a house.” These errors are more common in the later phases of L2 acquisition [17]. Misordering errors occur when there is a wrong placement of a morpheme or collection of morphemes in a well-formed utterance [17]. For example, in the sentence “we cannot anymore live on this planet”, the expression “anymore live on this planet” should be reordered to “live on this planet anymore”.

2.3. Previous Studies on Error Analysis in Spoken English

Most of the studies on error analysis of learners of English are based on written materials [9,22]. Compared to written data, error analysis of spoken data is more challenging because it requires transcription. Error analysis of speaking data can be broadly classified into two types: (a) classification of errors according to grammatical elements and/or (b) classification of errors according to the surface strategy taxonomy (e.g., omission, addition, misformation, and misordering; Dulay, Burt and Krashen [17]).

There have been several studies that used the classification of errors made by Korean learners of English based on grammatical categories. Chin [23] investigated the types of spoken errors produced by low-level Korean learners of English. Her study revealed that errors related to the use of noun phrases and verb phrases were most frequently made in picture description tasks. Among noun phrase-related errors, article use errors were most common, followed by noun form errors. Kim [24] analyzed the types of spoken errors made by Korean college students majoring in secretarial studies in English-speaking tests using simulated telephone conversations. The study identified six major error categories, with the most frequent errors being noun phrase errors, followed by verb phrase errors, clause/sentence structure errors, adjective phrase errors, and prepositional phrase errors. Among the subcategories of errors, errors in the use of articles were most frequently found, followed by errors in word order, how to address others, verb tense choice, subject or verb omission, and *BE-verb* misuse. Kim, et al. [25] investigated spoken errors made by Korean high school students, examining how errors vary by English speaking proficiency level. The study found a high frequency of errors related to verb omission, verb addition, verb choice, subject-verb agreement, and the use of *-ing* across all levels of learners. However, the use of the *-ing* form differed between high-proficiency-level students (e.g., use of the *-ing* form where unnecessary) and low-proficiency-level students (e.g., omitting *-ing* where necessary). Ahn [26] studied the types and frequency of grammar errors made by Korean college students while communicating with their peers in English and the impact these errors had on communication breakdown. She adopted the grammatical error categories used in Chin [23] and Kim [24], and found that errors related to noun phrases and verb phrases were the most prevalent. The frequency hierarchy of error subcategory was relevant to articles, verbs, prepositions omitted, word misuse, and pronouns, in this order. Conversation analysis showed that grammar errors rarely caused communication problems.

Noh [27] and Yoon [21] analyzed the spoken errors produced by Korean college students in speaking tests using a surface strategy classification: addition, omission, misformation, and misordering errors [17]. Noh [27] reviewed errors in the speech of L2 English learners to identify error types and causes, including unnatural expression errors and communication disorder errors, using the surface strategy classification. This study found that semantic errors were the most prominent, followed by grammatical errors such as in the choice of prepositions and parts of speech. Yoon [21] investigated and classified the

grammatical errors made by Korean EFL students in the TOEIC speaking test according to the surface strategy classification. She found that omission errors were the most prevalent (74.9%), followed by misformation (19.9%) and misordering (1.7%).

Several studies have used both grammatical elements and the surface strategy taxonomy as the main criteria for error classification. Back [28] analyzed preposition errors made by Korean EFL learners in speaking and writing, as well as the factors that caused these errors. The results showed that omission (about 40%) was the most prevalent error in the speaking corpus, followed by misformation (34%) and addition errors. In the case of preposition errors, on the other hand, addition (28.3%) was the most salient error, followed by omission (19.7%) and misformation (10.1%). Son and Chang [29] conducted a comprehensive analysis of the grammatical errors exhibited by Korean university students during their English speaking activities. Moreover, they examined the variations in error patterns based on the students' levels of proficiency. In the study, participants were asked to describe a picture and respond to a speaking prompt. The study categorized the grammatical errors into six groups with 45 subgroups. Overall, the results found that noun and verb phrases had the highest error rates. Grammatical errors in noun phrases were the most frequent, followed by verb phrases, preposition phrases, sentence structure, and adjectives. The error analysis by proficiency level showed that beginner-level learners made more grammatical errors overall than intermediate learners. More specifically, beginners had the most subject-verb agreement errors, while intermediate learners had more awkward expressions. Choi [9] analyzed the English speaking errors and error perceptions of Korean learners. The study used college students' TOEIC speaking test responses and transcribed their audio clips. The transcripts were then coded with the six grammatical errors (e.g., noun phrase, verb phrase, number/price unit, clause/sentence structure, prepositional phrase, adjective phrase, others) introduced in Chin [23] and Kim [24]. These six categories were subdivided into omission, addition, misformation, and lexical errors as proposed by Dulay, Burt and Krashen [17], Noh [27], and Yoon [21].

Overall, the studies investigated various grammatical error types and their relative frequencies in the spoken English data of Korean learners of English. However, many of the studies focused primarily on error frequencies without considering the learners' proficiency levels. While Kim, Pae, Hwang and Rhee [25] and Son and Chang [29] took proficiency levels into account when analyzing error types and their frequencies, they did not investigate whether each error solely affected the proficiency level or if all errors collectively affected students' proficiency levels. Furthermore, these studies rarely considered how prompts (or inputs) affected the types and frequency of errors produced by learners.

2.4. Methods of Analysis

The quantitative data in the current study were analyzed using four statistical methods: Friedman's test for comparing the frequency of lexico-grammatical errors across prompt topics within groups, Wilcoxon signed-rank test for pairwise comparison of the topic groups, hierarchical regression analysis for analyzing the impact of error types on lexico-grammatical scores, and many-facet Rasch measurement analysis for calculating examinees' lexico-grammatical scores.

The Friedman test [30], a non-parametric equivalent to the analysis of variance (ANOVA), is a statistical test used to analyze the differences between multiple related groups or treatments based on ordinal data. It is often used when the data are not normally distributed or when the assumptions for parametric tests are not met. The Wilcoxon signed-rank test [31], a non-parametric equivalent to the paired samples *t*-test, is commonly used when analyzing matched samples based on ordinal data. Hierarchical regression analysis is a statistical method used to examine the association between one dependent variable and multiple independent variables [32]. It is a type of multiple regression analysis in which independent variables are sequentially entered into the regression model based on a theoretical rationale, thus allowing statistical model comparison.

Many-facet Rasch measurement (MFRM) is an expanded form of the Rasch model designed to calibrate multiple variables, known as facets, that can influence assessment outcomes [33]. While the original Rasch model, introduced by Rasch [34], focuses on the relationship between two facets (examinee ability and item difficulty) to determine the probability of a correct answer, MFRM incorporates more than two facets simultaneously, such as examinee ability, item difficulty, rater severity, and test form [35]. MFRM provides valuable insights into the interplay among these facets, allowing for the estimation of their relative positions on a linear scale. In MFRM models, latent variables like examinee ability and item difficulty are expressed on the logit scale, which represents the natural logarithm of the odds ratio and spans from negative infinity to positive infinity [35]. These variables are assumed to exist on the same continuum, known as the item–person logit scale. Examinees positioned at the same point on this scale have approximately a 50% chance of correctly answering dichotomous items or achieving a specific item score on items with continuous rating scales [35,36].

3. Method

This section describes the methods employed in the present study. It commences by explaining the data source and the instruments used. The procedures for data collection and coding are then described in detail. The section concludes with a description of the data analysis techniques employed to address the research questions.

3.1. Data

The data were collected from an oral proficiency interview administered to 32 Korean college students in Korea. The examinees selected for the interview were required to have a score between 600 and 900 on the Test of English for International Communication (TOEIC) or an equivalent proficiency approved by the researchers. The speech of the examinees was transcribed for further analysis and the lengths of their utterances varied from 8 to 305 tokens across different prompt topics ($Mean = 152$, $SD = 65$), as presented in Table 1.

Table 1. Description of examinees' utterance lengths (tokens).

Topics	Min	Max	Mean	SD
Q1. School bag	45	303	147	59
Q2. Online class	50	305	170	68
Q3. Free housing	8	268	140	65
Total	8	305	152	65

Note: SD stands for standard deviation.

3.2. Oral Proficiency Interview

The oral proficiency interview (OPI) used in the current study assesses the effectiveness of college students' communication in English in university life or classroom situations. The OPI consists of three questions: one descriptive and explanatory speaking question, one compare-and-contrast question, and one hypothetical question as illustrated in Table 2.

Table 2. Interview prompts for the study.

No	Prompt Characteristics	Prompts
1	Descriptive Present tense	What items do you usually have in your school bag? Please describe them.
2	Compare-and-contrast Present tense	What are some benefits and drawbacks of taking online classes from home?
3	Hypothetical Conditionals	How would the world change if everyone had free housing?

The interview typically takes approximately ten minutes to complete. Prior to the three main questions, the interview begins with a one-minute warm-up conversation between the interviewer and the examinee. The oral production of the examinee during the warm-up is not graded. Each question of the interview requires the examinee to respond for two minutes without any preparation time. If examinees' responses are shorter than two minutes, the interviewer asks follow-up questions to help the examinees continue producing enough speaking samples.

Two raters assign scores on a 13-point scale with four proficiency levels as shown in Table 3. The current rating scale for lexico-grammar was adapted from Cotos' [37] rubric, which was originally designed to measure organization, fluency, lexico-grammar, and pronunciation.

Table 3. Lexico-grammatical scoring rubric.

Level	Score	Description
Excellent	13	Native-like with sophisticated, appropriate, precise vocabulary and grammar
Very Strong	11–12	Rich vocabulary and accurate grammar; a few unusual expressions or minor problems possible
Strong	9–10	Adequate, but not stellar vocabulary; occasional or slight problems with expression or lexical and grammatical forms
Adequate	7–8	Good vocabulary; good but inconsistent use of all time frames; noticeable but not serious problems with expression or lexical and grammatical forms
Limited	5–6	Fair vocabulary; fair use of time frames and expression; sentence structure problems
Very Limited	3–4	Limited vocabulary; lack of grammar control; frequent and/or serious errors
Poor	1–2	Poor to almost lacking control of vocabulary and grammar
Not competent	0	Lack of vocabulary and grammar

3.3. Raters and Coders

Two raters were recruited to score 96 audio recordings using a lexico-grammatical rating scale as outlined in Table 3. Both raters were doctoral students in applied linguistics at a U.S. university and had received training as OPI interviewers and raters. Additionally, they had extensive experience as English as a second language (ESL) instructors having taught ESL courses at a tertiary education institution in the U.S. for over three years. For lexico-grammatical error coding, a rater first reviewed the examinees' transcripts and identified the errors. Then, two other coders, who were non-native speakers of English and held linguistics-related college degrees, were invited to review the initial coding and count the frequency of each error. One of the researchers reviewed the error frequency coding and finalized it.

3.4. Coding Criteria

The error coding criteria were designed following the surface strategy classification (i.e., addition, omission, misformation, and misordering errors) proposed by Dulay, Burt and Krashen [17], as well as the error groups (i.e., noun phrase errors, verb phrase errors, adjective phrase errors, prepositional phrase errors, sentence structure errors, and miscellaneous errors) employed by Chin [23], as presented in Table 4. For example, the utterance “*I didn't go school” is missing the preposition *to*. This utterance is coded as a sentence with an “adjective phrase with omission error.” Omission, addition, and misformation errors can be found across all lexico-grammatical units, but misordering is specific to sentence structure.

Table 4. Lexico-grammatical error coding criteria and examples.

Grammatical Errors	Surface Strategy Taxonomy			
	Omission	Addition	Misformation	Misordering
Noun Phrase Errors	(my) economics book	very common things in (the) learning	my laptop (notebook)	n/a
Verb Phrase Errors	I stopped (working)	make my nephew (is) sick	many trees will be (are) gone	n/a
Adjective Phrase Errors	a (higher) percentage of getting As	we don't have to donate a (more the) house	it is very efficient (effort).	n/a
Prepositional Phrase Errors	I didn't go (to) school	socializing(in) online	would not agree with (about) it.	n/a
Sentence Structure	First, (it) is safe	houses is getting more (cost is) expensive	their cover(s) are (is) pretty beautiful.	nothing is bad (bad is nothing)
Miscellaneous	the skill for science, (and) art.	because (if) they have home to live in.	So not (no) tired.	n/a

Note: () indicates errors. n/a indicates that there is no corresponding example utterance.

3.5. Procedures

This study was conducted using the following procedures. First, the interviewer, who was also one of the raters, interviewed the examinees using the three given prompts in Table 2. Each examinee was asked to respond to each question for approximately two minutes. The interview was conducted and recorded via the online platform Zoom. Second, the examinee's performance was evaluated based solely on the audio recordings. Two raters used the lexico-grammatical evaluation rubric in Table 3 to score the examinees' responses. In order to minimize measurement errors in the rating, a rating session was prepared and the rating order was randomized. Rater training involved using ten sample audio clips to calibrate the rating scales among the raters before the rating session. The rating order of audio clips was randomized for each rater to mitigate the effects of the rating order. Third, all audio recordings were transcribed and one of the raters reviewed the transcription for lexico-grammatical errors. Fourth, all lexico-grammatical errors were coded for statistical analysis by the coders based on the rater's lexico-grammatical review. Finally, the lexico-grammatical errors and the rating scores were statistically analyzed using the Friedman test, which is a non-parametric alternative to ANOVA; the Wilcoxon signed-rank test, which is a non-parametric alternative to the paired samples *t*-test; the many-facet Rasch measurement (MFRM) model; and hierarchical regression analysis.

3.6. Data Analysis

The current study employed a quantitative approach to examine the variability in examinees' lexico-grammatical performance based on the interview prompt. To investigate the effect of task topics on the frequency of lexico-grammatical errors (Research Question 1), multiple Friedman tests were conducted, followed by pairwise comparisons using the Wilcoxon signed-rank tests. Non-parametric tests were chosen, as the linguistic measures were frequency data obtained from the linguistic output and did not meet the assumptions of a parametric analysis. Since the linguistic measures were not conceptually related, multiple univariate analyses were conducted instead of a multivariate analysis [38]. To address the issue of multiple comparisons in the follow-up pairwise comparisons with the Wilcoxon signed-rank tests, the significance level was adjusted using a Bonferroni correction ($\alpha = 0.05/3 = 0.017$), considering that there were three topics [39].

To analyze the effect of interview topics and related errors on examinees' lexico-grammatical performance (Research Question 2), a rating scale model of MFRM was employed to calculate examinees' fair lexico-grammatical scores. A *fully-crossed* or *complete* design was used where each rater assessed all examinees on all items. This design ensures the highest precision for parameter estimation and avoids missing data in the MFRM model [33]. The rating design used in the present study is depicted in Figure 1, where

each tick mark, or “√”, represents the scores assigned by raters. The final *fair* score was estimated using the computer program FACETS version 3.80 [40].

Raters	Topics	Audio Recordings								
		1	2	3	4	...	93	94	95	96
1	Q1, Q2, Q3	√	√	√	√	...	√	√	√	√
2	Q1, Q2, Q3	√	√	√	√	...	√	√	√	√

Figure 1. Illustration of rating designs for three-facet MFRM.

The rating scale model for the lexico-grammatical scores can be expressed as shown in Equation (1). Additionally, a hierarchical regression analysis was conducted to determine the extent to which the frequency of grammatical errors in the utterance explains the lexico-grammatical scores.

$$\ln[P_{njfk}/P_{njfk-1}] = \theta_n - \alpha_j - \tau_k, \quad (1)$$

where

P_{njfk} = probability of test taker n receiving a rating of k from rater j ;

P_{njfk-1} = probability of test taker n receiving a rating of $k - 1$ from rater j ;

θ_n = speaking ability of examinee n ;

α_j = severity of rater j ;

τ_k = difficulty of scale category k relative to scale category $k - 1$.

4. Results

This section presents the results of the data analysis conducted to answer the research questions using the findings from the quantitative analysis described in the previous section. It begins with the frequency of lexico-grammatical errors by topic in the OPI, employing non-parametric statistical tests for group comparison. It then describes the results of the regression analysis, aimed at explaining the effect of lexico-grammatical errors on the examinees' lexico-grammatical scores.

4.1. Lexico-Grammatical Errors by Topic

The total raw frequencies of lexico-grammatical errors by topic were 457, 621, and 475 for Q1: *school bag*, Q2: *online class*, and Q3: *free housing*, respectively. As shown in Table 5, the number of noun phrase (NP) errors was 725 (46.7%), verb phrase (VP) errors were 289 (18.6%), sentence structure errors were 174 (11.2%), prepositional phrase (PP) errors were 136 (8.8%), adjective phrase (AP) errors were 54 (3.5%), and miscellaneous errors were 175 (11.3%). Noun phrase errors were the most prevalent, followed by verb phrase errors, sentence structure errors, prepositional phrase errors, and adjective phrase errors, except for miscellaneous errors, in every topic.

As the length of utterance varied, these error frequencies were normalized using the average length of utterance per topic in order to compare the frequency among topics. The average lengths of utterances per topic were 147, 170, and 140 for Q1: *school bag*, Q2: *online class*, and Q3: *free housing*, respectively, as shown in Table 1. Thus, the normalized frequency was calculated by dividing the raw frequency of each category by the average length of utterance (e.g., NP of Q1 can be calculated as $258/147 \times 100 = 176$), as shown in Table 6.

The normalized frequencies of total lexico-grammatical errors for Q1: *school bag*, Q2: *online class*, and Q3: *free housing* were 311, 365, and 313, respectively. These normalized frequencies were used for further analysis. As shown in Table 6, noun phrase (NP) errors were 465 (47.0%), verb phrase (VP) errors were 184 (18.6%), sentence structure errors were 110 (11.2%), prepositional phrase (PP) errors were 86 (8.7%), adjective phrase (AP) errors were 34 (3.4%), and miscellaneous errors were 110 (11.2%). Noun phrase errors were the most prevalent, followed by verb phrase errors, in every topic.

Table 5. Raw error frequency across topics.

Grammatical Errors	Topic			Total (%)
	Q1. School Bag (%)	Q2. Online Class (%)	Q3. Free Housing (%)	
NP	258 (56.5)	255 (41.1)	212 (44.6)	725 (46.7)
VP	61 (13.3)	114 (18.4)	114 (24.0)	289 (18.6)
AP	12 (2.6)	28 (4.5)	14 (2.9)	54 (3.5)
PP	30 (6.6)	65 (10.5)	41 (8.6)	136 (8.8)
Sentence Structure	55 (12.0)	76 (12.2)	43 (9.1)	174 (11.2)
Miscellaneous	41 (9.0)	83 (13.4)	51 (10.7)	175 (11.3)
Total	457 (100.0)	621 (100.0)	475 (100.0)	1553 (100.0)

Note: NP: noun phrase, VP: verb phrase, AP: adjective phrase, PP: prepositional phrase.

Table 6. Error frequency normalized by average length of utterances across topics.

Grammatical Errors	Topic			Total (%)
	Q1. School Bag (%)	Q2. Online Class (%)	Q3. Free Housing (%)	
NP	176 (56.5)	150 (41.1)	139 (44.6)	465 (47.0)
VP	41 (13.3)	67 (18.4)	75 (24.0)	184 (18.6)
AP	8 (2.6)	16 (4.5)	9 (2.9)	34 (3.4)
PP	20 (6.6)	38 (10.5)	27 (8.6)	86 (8.7)
Sentence Structure	37 (12.0)	45 (12.2)	28 (9.1)	110 (11.2)
Miscellaneous	28 (9.0)	49 (13.4)	34 (10.7)	110 (11.2)
Total	311 (100.0)	365 (100.0)	313 (100.0)	989 (100.0)

Note: the frequency is normalized by topic length. NP: noun phrase, VP: verb phrase, AP: adjective phrase, PP: prepositional phrase.

The results of the Friedman tests examining each lexico-grammatical error by prompt topic revealed significant differences in noun phrase, verb phrase, and prepositional phrase errors: noun phrase errors, $\chi^2(2) = 7.04$, $p = 0.030$ *; verb phrase errors, $\chi^2(2) = 11.24$, $p = 0.004$ **; adjective phrase errors, $\chi^2(2) = 3.98$, $p = 0.137$; prepositional phrase errors, $\chi^2(2) = 11.33$, $p = 0.003$ **; sentence structure errors, $\chi^2(2) = 0.807$, $p = 0.668$; and miscellaneous errors, $\chi^2(2) = 1.590$, $p = 0.452$. As the Friedman tests for noun phrase, verb phrase, and prepositional phrase errors were statistically significant, *post hoc* analyses were conducted using Wilcoxon signed-rank tests as shown in Table 7. The results of the Wilcoxon signed-rank tests showed that only the verb phrase errors between Q1 (school bag) and Q2 (online class) ($Z = -2.84$, $p = 0.005$, $r = -0.36$), between Q1 and Q3 (free housing) ($Z = -3.69$, $p = 0.000$, $r = -0.46$), and the prepositional phrase errors between Q1 and Q2 ($Z = -2.96$, $p = 0.003$, $r = -0.37$) were statistically significantly different.

Table 7. Wilcoxon signed-rank tests of noun, verb, and prepositional phrase errors ($df = 2$).

Categories	Z ^a	p ^b	r ^c	
NP	Q1 vs. Q2	-2.22	0.027	-0.28
	Q1 vs. Q3	-1.88	0.060	-0.24
	Q2 vs. Q3	-0.25	0.805	-0.03
VP	Q1 vs. Q2	-2.84	0.005 *	-0.36
	Q1 vs. Q3	-3.69	0.000 *	-0.46
	Q2 vs. Q3	-1.79	0.073	-0.22
PP	Q1 vs. Q2	-2.96	0.003 *	-0.37
	Q1 vs. Q3	-1.28	0.201	-0.16
	Q2 vs. Q3	-1.67	0.095	-0.21

Note: ^a Wilcoxon signed-rank test; ^b asympt. sig (2-tailed): * $p < 0.017$; ^c effect size. $r = 0.10$ is a small effect; $r = 0.30$ is a medium effect; and $r = 0.50$ is a large effect [41]. NP: noun phrase, VP: verb phrase, PP: prepositional phrase.

4.2. Lexico-Grammatical Errors and Lexico-Grammatical Scores

A correlation analysis was conducted to examine the relationship between the lexico-grammatical scores (dependent variable; DV) and the error variables, in order to identify potential independent variables (IVs) for a hierarchical regression analysis. Table 8 presents the correlations among the IVs and the DV. The Pearson correlations indicated that the lexico-grammatical scores were negatively correlated with noun phrase errors ($r = -0.36$, $p < 0.001$ ***), verb phrase errors ($r = -0.35$, $p < 0.001$ ***), adjective phrase errors ($r = -0.06$, $p = 0.546$), prepositional phrase errors ($r = -0.08$, $p < 0.461$), sentence structure errors ($r = -0.21$, $p < 0.05$ *), and miscellaneous errors ($r = -0.13$, $p = 0.193$). Additionally, positive correlations were found between the lexico-grammatical scores and the length of utterance in tokens ($r = 0.76$, $p < 0.001$ ***) and tokens per AS-unit ($r = 0.51$, $p < 0.001$ ***).

Table 8. Correlations between lexico-grammatical score and error variables.

	NP Error	VP Error	AP Error	PP Error	Sentence Structure Error	Miscellaneous Error	Utterance Length (Tokens)	Tokens per AS-Unit
Lexico-Grammatical Scores	-0.36 ***	-0.35 ***	-0.06	-0.08	-0.21 *	-0.13	0.76 ***	0.51 ***

Note: * $p < 0.05$, *** $p < 0.001$. NP: noun phrase, VP: verb phrase, AP: adjective phrase, PP: prepositional phrase.

A hierarchical regression analysis was conducted on the lexico-grammatical scores using four independent variables, as presented in Table 9. Because adjective phrase errors, prepositional phrase errors, and miscellaneous errors were not found to be statistically significant predictors of the lexico-grammatical scores, these variables were not included in the initial set of independent variables for the regression analysis. The hierarchical multiple regression analysis showed that in Model 1, noun phrase errors and verb phrase errors significantly contributed to the regression model, $F(2, 93) = 13.18$, $p < 0.001$ ***, explaining 22% of the variance in the lexico-grammatical scores, as depicted in Table 9. The addition of sentence structure to the regression model did not result in a statistically significant improvement, so it was not included in the final model. The potential impact of topic variation was also examined in Model 1, but it was not found to be a significant contributor to the regression model. Therefore, model comparison was performed without considering topic variation.

Table 9. Multiple regression analysis.

	Model	R	R ²	B	SE	β	t	p
1	(constant)			7.96	0.46		17.42	<0.001 ***
	NP Error	0.47	0.22	-0.23	0.07	-0.31	-3.39	0.001 **
	VP Error			-0.38	0.12	-0.30	-3.26	0.002 **
2	(constant)			2.72	0.60		4.56	<0.001 ***
	NP Error			-0.10	0.05	-0.13	-2.02	0.047 *
	VP Error	0.81	0.65	-0.23	0.08	-0.19	-2.94	0.004 **
	Utterance Length (Tokens)			0.02	0.00	0.60	8.20	<0.001 ***
	Tokens per AS-unit			0.10	0.04	0.16	2.26	0.026 *

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. B is unstandardized beta; SE is standard error; β is standardized beta. NP: noun phrase, VP: verb phrase.

In the final model, or Model 2, the variables of total utterance length and tokens per AS-unit, which represent the average length of each utterance, were added to the initial model, or Model 1. This final model accounted for 65% of the variance in the lexico-grammatical scores. The inclusion of the utterance length-related variables in the regression model explained an additional 43% of the variance ($\Delta R^2 = 0.65 - 0.22$) in the lexico-grammatical scores, yielding a significant overall improvement, $F(2, 91) = 11.05$, $p < 0.001$ ***. The

final model indicates that for each additional noun phrase error, the lexico-grammatical scores decreased by 0.10, while holding the effects of other independent variables constant. Similarly, for each additional verb phrase error, the lexico-grammatical scores decreased by 0.23, holding the effects of other variables constant. On the other hand, an increase of one token in the utterance length lead to a 0.02 increase in the lexico-grammatical scores, while controlling for the effects of other variables. Likewise, an increase of one token per AS-unit resulted in a 0.10 increase in the lexico-grammatical scores, while holding the effects of the other variables constant.

According to the regression analysis, the results suggest that reducing noun phrase errors by approximately ten can result in an increase of one level in the lexico-grammatical score. Similarly, reducing verb phrase errors by about five can lead to an increase of one-level in the lexico-grammatical score. In terms of utterance length, producing ten additional words in the response was associated with a 0.20 increase in the lexico-grammatical score. However, considering that the average utterance time was two minutes, this increase may not be considered significant in terms of score gain. Speaking one more word per AS-unit statistically significantly contributed to the variability in lexico-grammatical scores, even when the utterance length variable was already included in the model. This additional explanation may be attributed to the complexity of utterances, which was not explicitly included in the current model. It is worth noting that more complex utterances generally require more tokens per AS-unit, and complexity is often taken into account when evaluating lexico-grammatical scores in language assessment.

5. Discussion

This section discusses the results of the data analysis and provides a summary of the key findings in the current study. It further presents the implications of employing error analysis in the measurement of lexico-grammatical scores in the oral proficiency interview.

5.1. Lexico-Grammatical Error Frequency by Topic

Based on the results of the error analysis, the frequency of lexico-grammatical errors in the examinees' English speaking tasks was compared across different error categories. It was found that the most errors were made by the examinees in noun phrases, verb phrases, but also in sentence structure, prepositional phrases, and adjective phrases (excluding miscellaneous errors). These findings align with numerous previous studies on error analysis of Korean learners of English, which have consistently shown that Korean learners tend to make the most errors in noun and verb phrases [9,23,24,29]. As the first research question aimed to investigate whether Korean learners of English would exhibit different error rates based on the topic, a comparison of error rates was conducted while adjusting for the same overall utterance length. This was necessary as the examinees' utterance length varied depending on the topics.

The results indicated that the descriptive task (*Q1: school bag*) exhibited a higher frequency of noun phrase errors compared to the compare-and-contrast task (*Q2: online class*) and the hypothetical task (*Q3: free housing*). Generally, errors are anticipated to be more prevalent in more challenging tasks, such as the compare-and-contrast task and the hypothetical task [12,42]. However, in the current study, noun phrase errors were unexpectedly more frequent in the descriptive task. This unexpected result could be attributed to the required language used to answer each question. The first prompt demands that examinees use descriptive language related to the items they carry in their bags, which elicits more noun phrase expressions, thus leading to the creation of more noun phrase errors.

On the other hand, prepositional errors were found to be statistically more prevalent in the compare-and-contrast task than in the descriptive task. No statistical differences were observed between the descriptive task and the hypothetical task. This result may be attributed to the shorter utterance length typically observed in the hypothetical task. Given that hypothetical conditional questions are cognitively more demanding compared to descriptive or compare-and-contrast prompts, it is expected that learners would make

more errors in the hypothetical conditional task. However, the more challenging nature of the tasks could lead to shorter utterances and, consequently, lower chances of making errors. Additionally, examinees could have used avoidance strategies to minimize the occurrence of errors in response to the hypothetical questions [43]. Prompt topic, including task types, can affect ESL learners' behaviors [44] as well as language errors.

5.2. Effect of Lexico-Grammatical Errors on Lexico-Grammatical Scores by Topic

The findings of this study support the hypothesis that topic variation can influence L2 performance, as evidenced by the statistically different error frequencies observed across topics. However, the results also suggest that variations in error frequency do not necessarily result in differences in lexico-grammatical scores. Contrary to prevailing assumptions in L2 oral production tasks [45–47], this study's findings indicate that the task topic does not exhibit a statistically significant influence on lexico-grammatical scores. Plausibly, this lack of statistical significance can be attributed to the minimal or negligible effects of task topics on linguistic output and/or the subconscious adjustment made by raters to account for variations in lexico-grammatical features prompted by different task topics. Notably, in the present study, the latter explanation appears more reasonable, as evidenced by the statistical significance of noun phrase errors and verb phrase errors varying across task topics. Thus, it can be inferred that the impact of task topic on lexico-grammatical scores may be indirectly mediated through the presence of lexico-grammatical errors, as depicted in Figure 2.

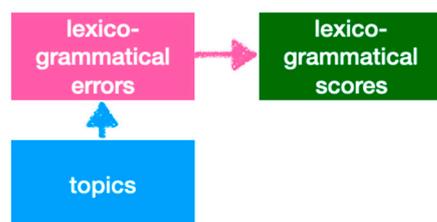


Figure 2. Plausible impact of task topic on lexico-grammatical scores.

The findings of this study also indicate that lexico-grammatical scores were affected by both lexico-grammatical errors and the length of utterances. To address the potential impact of utterance length on error frequency, the frequency of errors in each utterance was normalized by dividing it by the corresponding utterance length. Consequently, it is crucial to include utterance length as an independent variable in the regression model. According to the comparison of the regression models, Model 2, which includes the additional variable of utterance length, explains approximately 43% more variance in lexico-grammatical scores compared to Model 1, which only considers noun phrase errors and verb phrase errors. This increase in the explained variance exceeds the combined impact of noun phrase errors and verb phrase errors. These findings suggest that lexico-grammatical scores are not solely determined by the presence of specific lexico-grammatical errors. Instead, lexico-grammatical scores may encompass the use of a diverse range of vocabulary or expressions, which may be indirectly reflected in the utterance length variable. This interpretation implies that raters perceive longer utterances as indicative of a higher level of lexico-grammatical proficiency. Therefore, when interpreting lexico-grammatical errors, it is important to acknowledge that lexico-grammatical scores encompass both positively related features, such as vocabulary use, utterance length, and utterance complexity, as well as inversely related features, such as lexico-grammatical errors.

6. Conclusions and Implications

The purpose of this study was to identify the types of errors made by Korean college students when speaking English in different task contexts, and to investigate the impact of these errors on their scores for lexico-grammatical proficiency. This study found that noun phrase errors were the most frequent type of error, followed by verb phrase errors, across

all three tasks. These findings are partially consistent with previous research [23,24,29], which has also shown that noun phrase and verb phrase errors are the most frequent lexico-grammatical errors made by Korean college students. The frequency of verb phrase and prepositional phrase errors was lower in the descriptive task than in the compare-and-contrast and hypothetical conditioning tasks. This may be because descriptive tasks are less cognitively demanding than compare-and-contrast or hypothetical tasks. The task complexity has been found to affect task performance, as noted by Prabhu [48], Robinson [49], and Skehan [50].

This study has also demonstrated that lexico-grammatical scores were influenced by utterance length and tokens per AS-unit, while being negatively affected by noun phrase errors and verb phrase errors. Other studies have investigated the frequency of errors by examinees' proficiency levels [25,29], but did not explore how these errors affect proficiency scores. Chin [23] investigated the significance level of errors judged by native English speakers, but the significance level was not quantitatively verified. Her study has limitations, as it is challenging to apply the significance of errors to judging examinees' proficiency. Although all potential predictor variables of lexico-grammatical scores in the hierarchical regression model were not fully reviewed, the current study provides a general understanding of how errors impact lexico-grammatical scores while controlling for the length of utterances.

Pedagogical implications can be drawn from this study regarding the importance of error types in oral communication. Noun phrase errors and verb phrase errors are known to be the most common errors made by Korean ESL learners [23,24,29]. This study confirms that these error types, along with the length of the utterance, significantly affect lexico-grammatical scores in speaking tests. The study results suggest that teachers should prioritize addressing noun phrase and verb phrase errors, as these errors convey more semantic information than other error types, such as prepositional phrase errors. Additionally, as the frequency of error types varies depending on the question type, it is essential to teach oral communication in a way that considers task characteristics and contextual factors such as field, tenor, and mode.

This study has demonstrated some important findings regarding the lexico-grammatical errors made by English language learners, but there were some limitations. First, the study's classification of lexico-grammatical errors based on the surface strategy classification of Dulay, Burt and Krashen [17], and the comprehensive error groups of Chin [23], is not necessarily exhaustive. Second, the multiple regression model used in the study was based on a relatively small sample size. Therefore, some predictor variables might have been excluded from the model due to a lack of statistical power. Lastly, the 13-point scoring scale used in the study was adapted from the pre-existing OPI test. Given the size of the rating scale, some scores on the rubric may not have been used.

Further studies are needed to investigate how the frequency of lexico-grammatical features can be used to predict oral communication quality. This could be achieved by analyzing larger and more diverse samples of oral communication, taking into account variations in topic, field, and mode. Additionally, studies using more diverse statistical analyses of frequency data and a wide range of lexico-grammatical error features are required to provide more precise information about the frequency of different error types and their impact on scores in oral communication tests.

Author Contributions: Conceptualization, Y.W. and S.K.; methodology, Y.W.; formal analysis, Y.W. and S.K.; investigation, Y.W. and S.K.; resources, S.K.; writing—original draft preparation, Y.W.; writing—review and editing, S.K.; project administration, S.K.; funding acquisition, S.K. and Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A2A03067523).

Institutional Review Board Statement: The study was approved by the Institutional Review Board (or Ethics Committee) of Seoul National University (IRB No. 2202/003-005, February 16, 2022) for studies involving humans.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Numerical data will be available upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Corder, S.P. The significance of learner's errors. *IRAL Int. Rev. Appl. Linguist. Lang. Teach.* **1967**, *5*, 161–170. [\[CrossRef\]](#)
2. Brown, H.D.; Lee, H. *Teaching by Principles: An Interactive Approach to Language Pedagogy*, 4th ed.; Pearson: London, UK, 2015.
3. Foster, P.; Skehan, P. The influence of planning and task type on second language performance. *Stud. Second Lang. Acquis.* **1996**, *18*, 299–323. [\[CrossRef\]](#)
4. Lennon, P. Investigating fluency in EFL: A quantitative approach. *Lang. Learn.* **1990**, *40*, 387–417. [\[CrossRef\]](#)
5. Yoon, S.-Y.; Lu, X.; Zechner, K. Features measuring vocabulary and grammar. In *Automated Speaking Assessment: Using Language Technologies to Score Spontaneous Speech*; Zechner, K., Evanini, K., Eds.; Routledge: New York, NY, USA, 2020; pp. 123–137.
6. Kim, Y.J. A study on English listening strategies by Korean collegiate students. *Mirae J. Engl. Lang. Lit.* **2014**, *19*, 265–288.
7. Cheon, S.M. A study of effective instructional methods based on the error analysis of vocabulary from cyber university students' spoken learner corpus. *J. Foreign Stud.* **2017**, *39*, 167–192. [\[CrossRef\]](#)
8. James, C. *Errors in Language Learning and Use*; Longman: London, UK, 1998.
9. Choi, E.Y. Korean learner's English speaking errors and recognition of the errors. *Linguist. Assoc. Korea J.* **2015**, *23*, 129–152. [\[CrossRef\]](#)
10. Leaper, D.A.; Riazi, M. The influence of prompt on group oral tests. *Lang. Test.* **2014**, *31*, 177–204. [\[CrossRef\]](#)
11. Jennings, M.; Fox, J.; Graves, B.; Shohamy, E. The test-takers' choice: An investigation of the effect of topic on language-test performance. *Lang. Test.* **1999**, *16*, 426–456. [\[CrossRef\]](#)
12. Won, Y. The effect of task complexity on test-takers' performance in a performance-based L2 oral communication test for international teaching assistants. *J. Korea Engl. Educ. Soc.* **2020**, *19*, 27–52. [\[CrossRef\]](#)
13. Beaman, K. Coordination and subordination revisited: Syntactic complexity in spoken and written narrative discourse. In *Spoken and Written Language: Exploring Orality and Literacy*; Tannen, D., Ed.; Ablex: Norwood, NJ, USA, 1984; pp. 45–80.
14. Daller, H.; Xue, H. Lexical richness and the oral proficiency of Chinese EFL students. In *Modelling and Assessing Vocabulary Knowledge*; Daller, H., Milton, J., Treffers-Daller, J., Eds.; Cambridge University Press: Cambridge, UK, 2007.
15. Foster, P.; Tonkyn, A.; Wigglesworth, G. Measuring spoken language: A unit for all reasons. *Appl. Linguist.* **2000**, *21*, 354–375. [\[CrossRef\]](#)
16. James, C. *Errors in Language Learning and Use: Exploring Error Analysis*; Routledge: New York, NY, USA, 2013.
17. Dulay, H.C.; Burt, M.K.; Krashen, S.D. *Language Two*; Oxford University Press: New York, NY, USA, 1982.
18. Corder, S.P. *Error Analysis and Interlanguage*; Oxford University Press: Oxford, UK, 1981; Volume 2.
19. Schachter, J. An error in error analysis. *Lang. Learn.* **1974**, *24*, 205–214. [\[CrossRef\]](#)
20. Hammarberg, B. The insufficiency of error analysis. *Int. Rev. Appl. Linguist. Lang. Teach.* **1974**, *12*, 185–192. [\[CrossRef\]](#)
21. Yoon, H.-K. Grammar errors in Korean EFL learners' TOEIC speaking test. *Engl. Teach.* **2012**, *67*, 287–309. [\[CrossRef\]](#)
22. Fumero, K.; Wood, C. Verb errors in 5th-grade English learners' written responses: Relation to writing quality. *Languages* **2021**, *6*, 71. [\[CrossRef\]](#)
23. Chin, C. Error analysis: An investigation of spoken errors of Korean EFL learners. *Engl. Teach.* **2001**, *56*, 97–123.
24. Kim, Y.-S. Error analysis: A study of spoken errors of business English learners. *Korean Assoc. Sec. Stud.* **2004**, *13*, 179–200.
25. Kim, J.-S.; Pae, J.-K.; Hwang, P.-A.; Rhee, S.-C. Capturing the characteristics of Korean high school students at different proficiency levels: Analysis of verb errors in spoken English. *J. Korea Engl. Educ. Soc.* **2014**, *13*, 77–96. [\[CrossRef\]](#)
26. Ahn, S. The analysis of English speaking conversation of Korean college students: In relation to grammatical errors and their impact on the intelligibility. *New Korean J. Engl. Lang. Lit.* **2019**, *61*, 201–222. [\[CrossRef\]](#)
27. Noh, S. An Analysis of Errors by Steady-State L2 English Learners: Focused on Errors of Salient Unnaturalness and Communication Breakdown. Unpublished. Master's Thesis, Hanyang University, Seoul, Republic of Korea, 2010.
28. Back, J. Preposition Errors in Writing and Speaking by Korean EFL Learners. *Stud. Br. Am. Lang. Lit.* **2011**, *99*, 227–247.
29. Son, H.; Chang, W. An analysis of grammatical errors in English speaking of Korean university students. *Stud. Foreign Lang. Educ.* **2020**, *34*, 277–313. [\[CrossRef\]](#)
30. Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Am. Stat. Assoc.* **1937**, *32*, 675–701. [\[CrossRef\]](#)
31. Wilcoxon, F. Individual comparisons by ranking methods. *Biom. Bull.* **1945**, *1*, 80–83. [\[CrossRef\]](#)
32. Cohen, J.; Cohen, P.; West, S.G.; Aiken, L.S. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*; Routledge: New York, NY, USA, 2013.

33. Eckes, T. *Introduction to Many-Facet Rasch Measurement: Analyzing and Evaluating Rater-Mediated Assessments*, 2nd ed.; Peter Lang GmbH: Frankfurt am Main, Germany, 2015.
34. Rasch, G. *Probabilistic Models for Some Intelligence and Achievement Tests*; Danish Institute for Educational Research: Copenhagen, Denmark, 1960.
35. Bond, T.G.; Fox, C.M. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 3rd ed.; Routledge: New York, NY, USA, 2015.
36. McNamara, T.F. *Measuring Second Language Performance*; Longman: London, UK, 1996.
37. Cotos, E. *Oral English Certification Test (OECT): Rater Manual*; Iowa State University: Ames, IA, USA, 2014.
38. Huberty, C.J.; Morris, J.D. Multivariate analysis versus multiple univariate analyses. *Psychol. Bull.* **1989**, *105*, 301–307. [[CrossRef](#)]
39. Field, A. *Discovering Statistics Using SPSS*, 3rd ed.; Sage Publications: London, UK, 2009.
40. Linacre, J.M. A User's Guide to FACETS. 2014. Available online: <https://www.winsteps.com/a/Facets-ManualPDF.zip> (accessed on 20 May 2023).
41. Cohen, J. A power primer. *Psychol. Bull.* **1992**, *112*, 155–159. [[CrossRef](#)]
42. Robinson, P. Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *Int. Rev. Appl. Linguist. Lang. Teach.* **2005**, *43*, 1–32. [[CrossRef](#)]
43. Dörnyei, Z. On the teachability of communication strategies. *TESOL Q.* **1995**, *29*, 55–85. [[CrossRef](#)]
44. Borzova, E.; Shemanaeva, M. Interactive mobile home tasks vs. individual home tasks in university foreign language education at the upper-intermediate level. *Educ. Sci.* **2022**, *12*, 639. [[CrossRef](#)]
45. Gilabert, R. Effects of manipulating task complexity on self-repairs during L2 oral production. *Int. Rev. Appl. Linguist. Lang. Teach.* **2007**, *45*, 215–240. [[CrossRef](#)]
46. Ishikawa, T. The effect of task demands of intentional reasoning on L2 speech performance. *J. Asia TEFL* **2008**, *5*, 29–63.
47. Michel, M.C. Effects of task complexity and interaction on L2 performance. In *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*; Robinson, P., Ed.; John Benjamins: Amsterdam, The Netherlands, 2011; Volume 2, pp. 141–173.
48. Prabhu, N.S. *Second Language Pedagogy*; Oxford University Press: Oxford, UK, 1987.
49. Robinson, P. Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*; Robinson, P., Ed.; Cambridge University Press: Cambridge, UK, 2001; pp. 287–318.
50. Skehan, P. *A Cognitive Approach to Language Learning*; Oxford University Press: Boston, MA, USA, 1998.

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.