BUEFAD | 2022, Volume 11, Issue 3, 559-575
Bartın University Journal of Faculty of Education
dergipark.org.tr/buefad
DOI: 10.14686/buefad.1026153

Influence of Planning Time and Task Type on Lexis in L2 Oral Performance

İlkay GİLANLIOĞLU [a*]

a* Asst. Prof. Dr., Foreign Language Education Department, Faculty of Education, Eastern Mediterranean University, North Cyprus, (https://orcid.org/0000-0002-9902-3311) * ilkay.gilanlioglu@emu.edu.tr

Research Article	Received: **19.11.2021**	Revised: **14.12.2021**	Accepted: **28.03.2022**

ABSTRACT

Building up a stock of vocabulary with operational sufficiency is a formidable task for L2 learners. An adequate portion of their vocabulary needs to be readily accessible and retrievable for productive use in real communication. This study investigated the possible effects of the availability of planning time and task type on L2 learners' lexical production in oral performance. It employed a 2x2 experimental design, where there were two independent variables (planning time and task type), each with two levels (no planning and with planning; descriptive and narrative). Participants were 102 (51 females and 51 males) intermediate level Preparatory School university students at an English-medium university in Turkey. The 51 dyads performed oral descriptive and narrative tasks under no planning and with planning conditions. The oral production of the participants were analyzed using two major measures: lexical complexity, and lexical accuracy. Lexical complexity was measured by word length in syllables and by lexical richness/variation with sub-measures of type-token ratio, lexical word range, grammatical word range, lexical-to-grammatical ratio and lexical density. Lexical accuracy was measured by the number of error-free clauses. The results revealed that lexical use is predominantly determined by task type rather than planning time. Narrative tasks elicited more complex and richer vocabulary than descriptive tasks. Availability of planning time appeared to positively influence the accuracy of lexis used, but at the cost of lower degree of richness/variation. The results also indicated that a focus on lexis could be induced through task design, which fosters various aspects of L2 lexical use.

Keywords: Lexical accuracy, lexical complexity, lexical richness/variation, planning time, task type

İkinci Dil Sözlü Performansta Planlama Zamanı ve Görev Türünün Sözcük Kullanımına Etkisi

ÖZ

İşlevsel yeterliliğe sahip bir sözvarlığı oluşturmak ikinci dil öğrenenler için zorlu bir görevdir. Özellikle, gerçek iletişimde üretken kullanım için sözvarlığının yeterli bir bölümünün kolayca erişilebilir ve geri çağrılabilir olması gerekmektedir. Bu çalışma, planlama zamanının varlığı ve görev türünün ikinci dil öğrenenlerin sözlü performansta sözcüksel üretimi üzerindeki olası etkilerini araştırmıştır. Her biri iki seviyeli (planlamasız ve planlamalı; betimsel ve öyküleyici) iki bağımsız değişkenin (planlama zamanı ve görev türü) olduğu 2x2'lik bir deneysel tasarım kullanılmıştır. Katılımcılar, eğitim dili İngilizce olan Türkiye'deki bir üniversitenin Hazırlık Okulu'nda öğrenim gören orta düzey dil yeterliğine sahip 102 (51 kadın ve 51 erkek) öğrenciden oluşmuştur. Ellibir (51) çift planlamasız ve planlamalı koşullarıyla sözlü betimsel ve öyküleyici görevleri gerçekleştirmiştir. Katılımcıların sözlü üretimleri iki ana ölçü kullanılarak çözümlenmiştir: sözcüksel karmaşıklık ve sözcüksel doğruluk. Sözcüksel karmaşıklık sözcüklerin içerdiği hece sayısı ve sözcüksel zenginlik/değişkenlik ölçütleri kullanılarak ölçülmüştür. Sözcüksel zenginlik/değişkenlik, alt-ölçütleri olan tür-belirteç oranı, içeriksel sözcük aralığı, dilbilgisel sözcük aralığı, içeriksel sözcük-dilbilgisel sözcük oranı ve sözcük yoğunluğu ile ölçülürken, sözcüksel doğruluk hatasız tümcelerin sayısıyla ölçülmüştür. Sonuçlar, sözcüksel kullanımın planlama zamanından ziyade ağırlıklı olarak görev türü tarafından belirlendiğini ortaya koymuştur. Öyküleme gerektiren görevler betimleme gerektiren görevlerden daha karmaşık ve daha zengin sözcük kullanımına yol açmıştır. Planlama zamanının varlığı kullanılan sözcüklerin doğruluğunu olumlu yönde etkilediği görülmüştür ancak bu daha düşük derecede sözcüksel zenginlik ve çeşitlilik pahasına gerçekleşmiştir. Sonuçlar ayrıca ikinci dil sözcük kullanımının çeşitli yönlerini besleyen görev tasarımı yoluyla sözvarlığına odaklanmanın sağlanabileceğini göstermiştir.

Anahtar kelimeler: Görev türü, planlama zamanı, sözcüksel doğruluk, sözcüksel karmaşıklık, sözcüksel zenginlik/değişkenlik

To cite this article in APA Style:
Gilanlıoğlu, İ. (2022). Influence of planning time and task type on lexis in L2 oral performance. *Bartın University Journal of Faculty of Education, 11*(3), 559-575. https://doi.org/10.1016/buefad.1026153

1 | INTRODUCTION

Lexis and grammar are usually two challenging areas for L2 learners. Competence in both is needed for effective communication. Beginners, though, may depend more on simple and frequent vocabulary than grammar. For example, the utterance 'I book' produced by someone pointing at the bookshelf filled with books may help to achieve some degree of communication but it could still be interpreted in many different ways including: 'I want a book', 'I want that book', 'I've read that book', 'I know that book', 'I want to borrow that book', 'I want to see/look at that book', and so on. As a response to the question 'Which book?', the language user may not be able to elude 'specificity', which can be achieved through grammar, and feel the need to say something like 'the one with the red cover in the left corner'. In communication, lexis and grammar are complementary to each other.

In spoken discourse, the degree of reliance on lexis or grammar is relative to different factors such as the L2 learner's language competence and contextual clues or shared knowledge. Tourists not speaking the language they are visiting usually take dictionaries or phrase books with them. They demonstrate 'over-reliance' on lexis. Their language may be referred to as 'survival language'. On other occasions, what regulates reliance on lexis or grammar is the extent of shared knowledge. As Widdowson (1990) illustrates, when a surgeon utters the word 'scalpel!' on the operating theatre, s/he will be given the right tool without having to produce a grammatically well-formed request like 'Can I have a scalpel please?' because of the shared contextual knowledge. In this respect, lexis and grammar can be placed on a continuum regulated by context. Along this continuum, roughly between lexis and grammar, 'lexico-grammatical units' (Widdowson, 1990) exist. They have also been labelled as 'lexicalised sentence stems' (Pawley & Syder, 1983) and 'lexical phrases' (Nattinger & DeCarrico, 1992). Research has shown that the use of lexical phrases improves the L2 learner's fluency (Derwing, 2017; Foster, 2020; Hobbs, 2005; Wood, 2001) and achieving communicative goals by means of such phrases results in self-confidence. Therefore, a good stock of vocabulary, whether it be single words, phrases or lexical phrases, has the potential to aid communication in the target language.

A major question then is how L2 learners can improve their vocabulary. Before tackling this question, it is useful to distinguish between receptive and productive vocabulary. These two types of vocabulary are interrelated with receptive (i.e., listening and reading) and productive skills (i.e., speaking and writing), respectively (Nation, 2001). "Receptive vocabulary use involves perceiving the form of a word while listening or reading and retrieving its meaning" whereas "productive vocabulary use involves wanting to express a meaning through speaking or writing and retrieving and producing the appropriate spoken or written form (Nation, 2001, pp. 24-25). Others have used a similar distinction – active/passive vocabulary (Meara, 1990; Corson, 1995). However, Meara (1990) sees the active-passive distinction in terms of different types of word association whereas Corson (1995) views it from the perspective of use. Though such a distinction has been problematized in the related literature (e.g. Melka Teichroew, 1982; Meara, 1997), it has informed particularly the development of vocabulary tests such as *Vocabulary Levels Test* (Schmitt et al. 2001; Webb, Sasao & Ballance, 2017) and *Yes/No Test* used for placement purposes (Meara & Miralpeix, 2017), and *word frequency* to objectively measure lexical sophistication in oral or written output (Crossley, Salsbury, McNamara & Jarvis, 2011; Laufer & Nation, 1995; Read, 2000). Schmitt (2010) argues that "[t]his dichotomy has great ecological validity, as virtually every language teacher will have experience of learners understanding lexical items when listening or reading, but not being able to produce those items in their speech or writing (p. 80)." This assertion also implies that productive use of vocabulary is more challenging for learners.

Interestingly, Webb (2009) found that productive learning facilitates both receptive and productive knowledge of vocabulary. Based on the findings of his study, he suggests that productive learning of vocabulary might be a more effective method. More recently, Uchihara & Clenton (2020) investigated the relationship between vocabulary size and second language speaking ability. They discovered that possessing a large vocabulary does not automatically lend itself to sophisticated lexical production in speech. As "… productive vocabulary use is moderated by the individual" (Uchihara & Clenton, 2020, p. 543), failure to produce lexically rich texts may not always be due to lack of vocabulary knowledge but rather to factors like lack of motivation and willingness to respond (Nation & Webb, 2011) and avoidance strategies (Skehan, 2009b). In the present study, considering the intricate nature of productive lexical use, it is hypothesized that engaging in actual language use in oral communication can help L2 learners develop their productive use of lexis. This is the kind of engagement with

meaningful language use channeled into lexical performance that forms the motivation for the study. The major research gap that the current study attempted to address is the absence of the lexical dimension of task performance. Skehan (2009a) reports that "[a] major area of omission concerns the lexical aspects of task performance" and that studies published in the past two decades have predominantly used "a restricted set of performance measures", namely, complexity, accuracy and fluency (CAF) (p. 107). Despite limited attempts such as the use of *lexical range* (Ortega, 1999), *type-token ratio* (Robinson, 2001), "in the main the lexical area has not been well served" (Skehan, 2009a, p. 107). Considering studies that involved the manipulation of task features (i.e., planning time and/or task type) in the last ten years or so, one can justifiably claim that Skehan's (2009a) assertion still holds as lexical measures employed are in scarcity alongside CAF (not necessarily all three aspects together) and mostly concerned with written (e.g., lexical complexity in EFL students' argumentative writing (Ong & Zhang, 2010); L2 writing lexical complexity (Johnson, Mercado & Acevedo, 2012; lexical complexity in collaborative L2 writing (Kang & Lee, 2019); lexical complexity in L2 writing as influenced by strategic planning and task structure (Tabari, 2020); lexical variety in L2 descriptive writing (Tabari, 2016)) rather than oral task performance (e.g., lexical diversity in native speakers' task performance (Foster & Tavakoli, 2009); lexical diversity as a sub-dimension of linguistic complexity in communicative adequacy (Revesz, Ekiert & Torgersen, 2016). Comparative studies concerned with both written and oral performances that use lexical measures are even rarer (e.g., Yu, 2009). In a more recent study Bui (2019) states that while the use of performance measures such as fluency, accuracy and complexity "is becoming a standard practice, lexical complexity as a distinctive area has received less attention in the task-based language teaching (TBLT) literature" (p. 1). In Bui's (2019) study, considered, in this respect, one of the recent exceptions involving several aspects of lexical use, lexical complexity is operationalized in three dimensions: lexical diversity, lexical density and lexical sophistication. Clearly, to date lexical measures have been overlooked in the relevant literature despite Skehan's (2009b) call for such action: "[It is] vital to incorporate some measure of lexis into task performance" (p. 512). Therefore, the present study aimed to deploy a variety of lexical measures to portray a fuller picture of lexical production in L2 oral performance.

THEORETICAL BACKGROUND

Two major theoretical perspectives have been advanced to explain second language acquisition (SLA), namely linguistic and cognitive approaches. Linguistic theorists claim that there is a language-specific module in the mind which manifests language acquisition. Linguistic theorists, also referred to as Universal Grammar (UG) theorists, have dwelled on linguistic competence, that is, the linguistic system underlying L2 grammars and their constructions. On the other hand, cognitive theorists do not view language as separate from other aspects of cognition. The mind is capable of processing all kinds of information, including linguistic information. They have been more concerned with knowledge (i.e., competence) and actual use (i.e., performance). The theoretical framework of the current study lies with the *Information Processing* view (Kahneman, 1973), which stems from the cognitive approach. According to this theory, human learning is not dependent on simply responding to stimuli but processing the input. In this view, language learning is seen as a complex cognitive skill. While the linguistic theorists referred to the mind using metaphors such as 'language-specific module' or 'language acquisition device' (LAD), cognitivists adopted the 'limited-capacity processor' metaphor. By analogy to computers, the mind works like a processor with limited capacity. As applied to L2 learning, L2 learners have limited attentional resources (Schmidt, 1995). In other words, there is a limit to what they can attend to at a given time. The two extensions of this theoretical position are the *Limited Attentional Capacity Model* (Skehan, 1998) and the *Cognition Hypothesis* (Robinson, 2001, 2003). The Limited Attentional Capacity Model proposes that the three dimensions of output quality – complexity, accuracy, fluency (CAF) – are in competition for attention and that there are trade-off effects, that is, learners achieve an increase in one or more at the cost of a decrease in another. While Skehan (1998) predicts reduced CAF as a result of increased task complexity, Robinson (2001, 2003, 2005) predicts enhanced complexity and accuracy, but reduced fluency. Although the two cognitive positions make similar predictions, they offer different theoretical explanations (for a more detailed comparative review, see Ong & Zhang, 2010). Subscribing to the Limited Attentional Capacity Model (Skehan, 1998), in order to increase the so-called 'limited attentional capacity', the L2 learner needs to free up some attentional resources. In doing so, it is hypothesized that practice, in the sense of actual language use, plays an important role. Practice can lead to L2 acquisition through *automaticity* and *restructuring* of the cognitive processes built in communicative tasks (e.g., DeKeyser, 2001; McLaughlin, 1990; for an extensive overview on the concepts, see Segalowitz, 2003). As more language

becomes automatic, the processor can attend to other new items in the input. Once noticed and taken in (Schmidt, 1990, 1992, 1995), the new items are likely to cause some degree of restructuring in the L2 learner's interlanguage. Once the restructuring is confirmed and reconfirmed through language use on many encounters it leads to further automaticity. The presence or absence of opportunities to use the language may account for some variability in learner language.

It follows then that automaticity is the ultimate goal in L2 learning and teaching. L2 learners are expected to reach a proficiency level where they can use language rather effortlessly, not having to pay attention to everything they want to say in a particular context. The process view of lexis, which the current study adopts, involves contextual manifestations of lexis to pave the way to automaticity. In this view, cognitive processes such as noticing, intake, restructuring and automaticity are to be fostered. Ultimately, the learning of lexis would take place through using lexis in discourse wherein those cognitive processes operate. To put it in perspective, the present study investigates the effects of pre-task planning and task type on lexical use in spoken discourse from an information processing viewpoint.

The regulation of the construct 'planning' has been the focus of a number of studies. One of the earliest studies compared *planned discourse* to *unplanned discourse* (Ochs, 1979). Various forms of planning have been investigated so far: micro-planning versus macro-planning (co-planning) (Crookes, 1989); *on-line planning* (Yuan & Ellis, 2003); and *pre-planning* or *pre-task planning* (Crookes, 1989; Ellis, 1987; Foster & Skehan, 1996, 1999; Wigglesworth, 1997; Mehnert, 1998; Ortega, 1999). Pre-task planning operationalized in these studies refers specifically to 'strategic planning' where the learner is given time to plan the content and language to use prior to the actual task performance (Ellis, 2005). As far as the length of pre-task planning is concerned, most studies allowed 10 minutes prior to the task (e.g., Crookes, 1989; Foster & Skehan, 1996, 1999; Ortega, 1999). However, different lengths of pre-task planning time (i.e., 1 minute, 5 minutes, 10 minutes) were also investigated (Mehnert, 1998). The general findings regarding planning time have revealed gains in complexity and fluency, but mixed results for accuracy.

The regulation of 'task type', on the other hand, has appeared fruitful in terms of language production. Many different task types have been examined: concrete/immediate tasks versus abstract/remote tasks (Foster & Skehan, 1996; Skehan & Foster, 1997); convergent versus divergent tasks (Duff, 1986; Pica et al. 1993); story-retelling (Ortega, 1999). The general findings with respect to task type show that concrete tasks (as opposed to abstract tasks) tend to reduce the information processing load, therefore they result in increased accuracy and fluency; however, the evidence for fluency is mixed. Divergent tasks (as opposed to convergent tasks) lead to more complex language output. Story-retelling induces increased fluency and linguistic complexity.

Thanks to their potential benefits in L2 acquisition, pre-task planning time and task type are viable constructs whereby spoken output of L2 learners can be wielded. Specifically, the current study looked into how the regulation of pre-task planning time and task type could generate manifestations of lexical use in oral performance. Such an investigation could be considered an attempt to depict one 'neglected' aspect of oral task performance – lexical use – through a number of lexical measures. The main research question that the study aimed to address was: Can a focus on lexis be induced through the regulation of task features – pre-task planning and task type? It was hypothesized that pre-task planning and task type would lead to more complex and more accurate lexical use in oral production. Lexical complexity was measured by syllabic range (i.e. monosyllabic word range, disyllabic word range, polysyllabic word range) and lexical variation/richness (i.e., type-token ratio, lexical word range, grammatical word range, lexical-to-grammatical ratio, lexical density); and lexical accuracy was measured by the number of clauses with lexical errors. The specific hypotheses are as follows:

Hypothesis 1: Monosyllabic word range will be wider in planned than unplanned conditions, as well as in narratives than descriptives.

Hypothesis 2: Disyllabic word range will be wider in planned than unplanned conditions, as well as in narratives than descriptives.

Hypothesis 3: Polysyllabic word range will be wider in planned than unplanned conditions, as well as in narratives than descriptives.

Hypothesis 4: Type-token ratio will be greater in planned than unplanned conditions, as well as in narratives than descriptives.

Hypothesis 5: Lexical word range will be wider in planned and unplanned conditions, as well as in narratives than descriptives.

Hypothesis 6: Grammatical word range will be wider in planned than unplanned conditions, as well as in descriptives than narratives.

Hypothesis 7: Lexical to grammatical ratio will be higher in planned than unplanned conditions, as well as in narratives than descriptives.

Hypothesis 8: Lexical density will be higher in planned than unplanned conditions, as well as in narratives than descriptives.

Hypothesis 9: Lexical accuracy (measured by a lower percentage of error-free clauses) will be greater in planned than unplanned conditions, as well as in narratives than descriptives.

The hypotheses are concerned with mainly two aspects of lexical use: *lexical complexity* (syllabic range: Hypotheses 1-3) and (lexical richness/variation: Hypotheses 4-8), and *lexical accuracy* (Hypothesis 9).

2 | METHOD

RESEARCH DESIGN AND DATA COLLECTION

The study employed an experimental design. There were two independent variables: pre-task planning time and task type. Pre-task planning time was operationalized in two conditions, i.e., no planning (no planning time provided prior to the task) and with planning (10 minutes planning time provided prior to the task). Task type was operationalized in two: descriptive and narrative. The descriptive task, based on the description and sorting out of a series of pictures, can be described as dialogic whereas the narrative task, based on telling a story from a set of sequenced pictures, can be described as monologic. Both task types consisted of two parallel tasks (i.e. two parallel descriptive tasks and two parallel narrative tasks) which had been piloted tested prior to the experiment. On the descriptive task, the pair of participants is supposed to interact with each other while describing pictures, and therefore produce a dialogue. The participants need to put the pictures in the right order by asking and answering questions, agreeing or disagreeing, doing confirmation checks, and so on. On the narrative task, though there is a pair (i.e., the speaker and the listener) only the speaker is supposed to tell the story. It is, therefore, a non-interactive monologue. The presence of a listener is to authenticate the task.

The research design is illustrated in Figure 1 below:

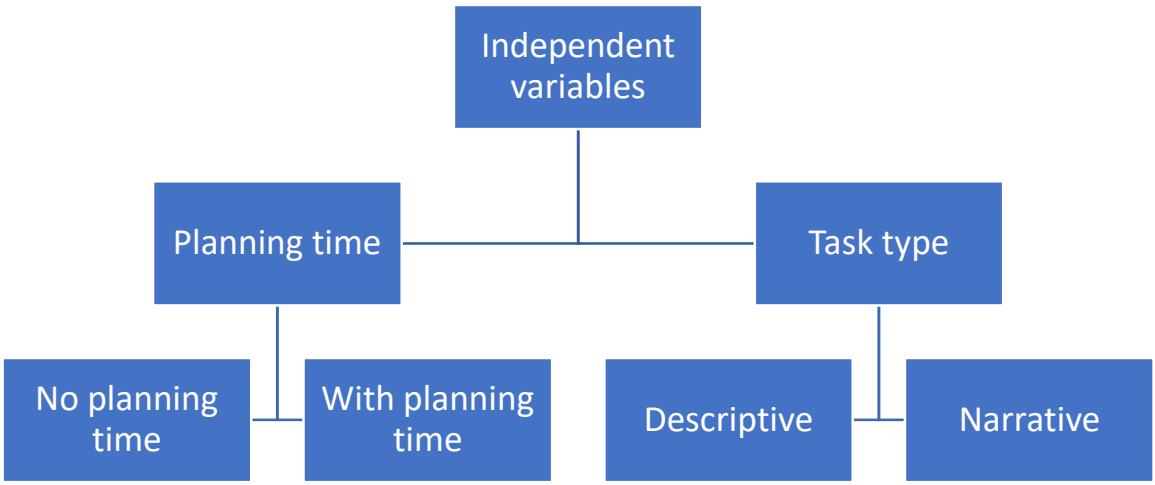

Figure 1. *Independent variables in the research design*

The 2-by-2 (2x2) experiment contained four experimental groups: -Descriptive, +Descriptive, -Narrative and +Narrative. Minus (-) represents no planning time while plus (+) represents with planning time conditions. Table 1 below outlines the experimental groups in the study:

Table 1. Experimental Groups: Planning by Task Type

	Descriptive	Narrative	Total
-Planning time	24 dyads	27 dyads	51 dyads
+Planning time	27 dyads	24 dyads	51 dyads
Total	51 dyads	51 dyads	

In -Descriptive and +Descriptive conditions, 24 and 27 dyads participated, respectively. In -Narrative and +Narrative conditions, 27 and 24 dyads took part, respectively. A total of 51 dyads participated in each of the planning time conditions (i.e., no planning and with planning).

PARTICIPANTS

The participants were a total of 102 intermediate level university students studying at the Preparatory School of a well-established English-medium university in Turkey. In terms of gender, there were equal numbers of males (N=51) and females (N=51). The average age was 18.

LEXICAL MEASURES AND RELIABILITY OF CODING

Two major lexical measures were used in the study: *lexical complexity* and *lexical accuracy*. In a review article, Suzuki (2017) reviews 40 studies on pre-task planning and reports that less than half of those studies used a lexical complexity measure and that it was generally operationalized as "the variety of word types available in the spoken production" (p. 21). Similarly, Ellis (2009) states as a result of his literature review that "[l]exical complexity was measured by means of type-token ratio and the number of different word types" (p. 495). In the current study, lexical complexity involved a variety of sub-measures, namely, type-token ratio, lexical word range, grammatical word range, lexical-to-grammatical ratio, and lexical density. Over one third of the data were coded by the researcher and a native-speaker instructor of English, who had been trained for that purpose. Reliability scores of codings are reported below as applicable.

Syllabic range, a new measure operationalized in this research study, was defined as the range of syllables in the participants' oral output. To date, no research concerned with 'pre-task planning' has used such a measure of lexical complexity. It was hypothesized that syllabic range was associated with phonological complexity. That is, the greater number of syllables a word has, the more phonologically complex it is. Words composed of multiple syllables (and indeed in English "… a vast majority of words are multi-syllabic" (Hamada, 2017, p. 1101)) were reasoned to be phonologically more complex in terms of processing. Studies in word-formation have drawn a parallel between the number of affixes in a word and its complexity (Marslen-Wilson et al., 1994). For instance, *un+able* and *un+ambigu+ous+ness* are complex words, the latter being more complex but less frequent while the former being less complex but more frequent (Nation, 2001, pp. 320-321). The examples refer to morphological complexity; however, the focus of the study is on phonological complexity, thus syllables, rather than morphemes, were counted. Phonological complexity is associated with word length, but not necessarily with the number of morphemes. Levelt, Roelofs & Meyer (1999) proposed that spoken word production is a complex and extremely fast process which entails processing at multiple levels: "After a first stage of *conceptual preparation*, word generation proceeds through *lexical selection*, *morphological and phonological encoding*, *phonetic encoding*, and *articulation* itself [my emphasis]" (p. 1). In this theory, after transition from conceptual/syntactic domain to phonological/articulatory domain, phonological encoding draws on syllabification (i.e., from phonological word to phonological syllables) which then leads to phonetic encoding that involves the activation of phonetic syllable scores in the syllabary by phonological syllables (Levelt et al. 1999). While some studies found that word length was not a significant variable in vocabulary learning (e.g., Rodgers, 1969), others argued otherwise (e.g., Phillips, 1981; Stock, 1976). Though linking word length to word difficulty may be problematic (Laufer, 1997), Coles (1982) found that word length had a strong influence on the success of non-native speakers of English in

recognizing written forms of English. Singleton (1999, p. 141) suggests two methodological problems that account for the diverse evidence of word length: "(1) word length can be variously calculated – in phonemes, graphemes, syllables or morphemes – and (2) it is difficult to disentangle length from other variables – notably morphological complexity." He further argues, in conformity to Levelt et al.'s (1999) speech model, that word-formation rules operate in correlation with phonological rules (Singleton, 2000). More recent work appears to be in favour of word-length measured in number of syllables. It has shown that "… the number of syllables positively correlated with word recognition time, suggesting that more syllables take longer to recognize" (Yap & Baloto, 2009; Perry et al., 2010 cited in Hamada, 2017). More specifically, regarding phonological processing Goldrick (2014) found that the syllabic dimension of phonological structure, along with the segmental and metrical dimensions, "are independently represented and retrieved" (p. 228). Consequently, it can be claimed that theory of speech production and relevant research evidence amply justify the use of syllabic range as an indicator of lexical complexity in oral performance.

All words were decomposed into their component syllables. A coder (a non-native instructor of English) was trained to divide the words into syllables. On the same one third of the data coded by the coder and the researcher, the intercoder reliability was 98%. The high level of reliability was due to using the same reference – Cambridge International Dictionary of English (1995) to look up most of the words for their component syllables, except perhaps those one-syllable words such as 'but', 'and', 'she'. The words were filed into three folders: monosyllabic (words composed of one syllable), disyllabic (words composed of two syllables) and polysyllabic (words composed of three or more syllables) words.

Type-token ratio, as a measure of lexical range, was calculated by dividing the number of different words (i.e., types) by the total number of words (i.e., tokens), following Ortega (1999), Ure (1971), and others.

Lexical word range was calculated applying the formula of types of lexical words (i.e., content words) divided by the total number of lexical words.

Grammatical word range was calculated by dividing the types of grammatical words (i.e. grammar words) by the total number of grammatical words.

Lexical-to-grammatical ratio was calculated by dividing the number of lexical words by the number of grammatical words. All closed-class functional words such as prepositions, conjunctions, articles, demonstratives, numerals, and the negation particle ('no') were counted as grammatical, and content words, i.e., all nouns, verbs, adjectives, and adverbs, as lexical (L. Ortega, personal communication, July 7, 1997). Those words not fitting either of these categories were grouped under the category of 'Other', and were excluded from the counts.

Lexical density was defined as the percentage of content words in the oral performance and calculated by dividing the number lexical words by the number of tokens and multiplying the result by one hundred.

Lexical accuracy was measured by the percentage of lexical choice errors and calculated following the formula of the number of lexical choice errors multiplied by one hundred and divided by the total number of clauses. Lexical choice errors were defined as "errors in lexical choice affecting words, phrases, or collocations" (Mehnert, 1998, p. 91). Repeated lexical choice errors were counted only once. Intercoder reliability for lexical choice errors was established at 95%.

DATA ANALYSIS

The lexical measures referred to in the hypotheses can be categorized as follows:

Table 2. Lexical Measures

Lexical complexity		Lexical accuracy
Syllabic word range	Lexical richness/variation	Number of clauses
Monosyllabic word range	Type-token ratio	with lexical errors
Disyllabic word range	Lexical word range	
Polysyllabic word range	Grammatical word range	
	Lexical-to-grammatical ratio	
	Lexical density	

As shown in Table 2 above, there were two major lexical measures – lexical complexity and lexical accuracy. Lexical complexity was operationalized in two sub-measures: syllabic word range and lexical richness/variation. Each category involved various aspects of lexical use. A multiple-measure approach employed in the study aimed to investigate the effects of plannedness and task type on lexical use in L2 oral performance.

The recorded speech was first transcribed and then coded for the measures used. Later the data were quantified using a specially designed computer program which was implemented in C++ using Microsoft Visual Studio 6.0. The computer program basically computed the occurrence of certain aspects of language use previously coded (e.g., grammar word vs. lexical/content words) and subjected the total counts to a set of formulas embedded in the program. This procedure was repeated for both planning conditions and task types. For instance, 'lexical density' is the ratio of lexical items to the total number of words in a text. To calculate lexical density lexical words need to be identified and counted. Similar word counts and calculations were performed on all the measures presented in Table 2 above. The resulting numerical data were then analyzed using the *Statistical Package for the Social Sciences (SPSS)* software to see whether the hypotheses were confirmed. For this purpose, two-way ANOVA was performed for each dependent variable with factors (i.e., pre-task planning time and task type) to determine the possible effect of planning, task type as well as the interaction effect.

RESEARCH ETHICS

The participants volunteered to take part in the study and their written informed consent was obtained prior to the implementation. The dialogues and monologues produced under the designated conditions during the experiment were recorded with the permission of the participants. Confidentiality and anonymity were ensured while collecting, storing, analyzing, and reporting the data.

3 | FINDINGS

In this section, the statistical results of the study for each measure are presented in reference to the related hypothesis/hypotheses. The hypotheses are grouped together as applicable under the relevant measures for coherence.

LEXICAL COMPLEXITY: SYLLABIC RANGE

Lexical complexity was measured in monosyllabic, disyllabic and polysyllabic word ranges. The three related hypotheses were collapsed into one as follows:

Hypothesis 1-3: Monosyllabic, disyllabic and polysyllabic word ranges will be wider in planned than unplanned conditions, as well as in narratives than descriptives.

Tables 3-5 present the results on the effects of planning and task type on lexical complexity in syllable ranges:

Table 3. Monosyllabic Word Range

	Descriptive			Narrative			Total		
	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N
-Planning	.345	.078	24	.465	.074	27	.408	.096	51
+Planning	.338	.102	27	.439	.070	24	.386	.101	51
Total	.342	.091	51	.453	.077	51	.398	.099	102

(Planning F = 1.01, p = .317; Task type F = 45.38, p = .000; Interaction F = .362, p = .549)

Table 4. Disyllabic Word Range

	Descriptive			Narrative			Total		
	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N
-Planning	.485	.125	24	.800	.091	27	.652	.191	51
+Planning	.443	.125	27	.762	.097	24	.593	.196	51
Total	.463	.125	51	.782	.095	51	.623	.195	102

(Planning F = 3.36, p = .070; Task type F = 209.17, p = .000; Interaction F = .010, p = .922)

Table 5. Polysyllabic Word Range

	Descriptive			Narrative			Total		
	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N
-Planning	.699	.255	24	.849	.171	27	.778	.225	51
+Planning	.659	.302	27	.745	.154	24	.700	.246	51
Total	.678	.279	51	.800	.170	51	.739	.238	102

(Planning F = 2.49, p = .118; Task type F = 6.68, p = .011; Interaction F = .501, p = .481)

Concerning the effect of planning, there were no statistically significant differences among syllabic ranges as computed for planned and unplanned conditions, monosyllabic word range (p=.317), disyllabic word range (p=.070), and polysyllabic word range (p=.118), respectively. Interestingly though, there were gains in no planning condition rather than planning condition. With respect to task type, there were statistically significant differences for all syllabic ranges – monosyllabic word range (p=.000), disyllabic word range (p=.000), and polysyllabic word range (p=.011), respectively. Therefore, each of Hypotheses 1-3 was partly confirmed.

LEXICAL COMPLEXITY: LEXICAL VARIATION/RICHNESS

Lexical richness was measured by the following sub-measures: type-token ratio, lexical word range, grammatical word range, lexical-to-grammatical ratio, lexical density. The two-way ANOVA results pertaining to each sub-measure are presented in Tables 6-10 below:

Table 6. Type-token Ratio

	Descriptive			Narrative			Total		
	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N
-Planning	.325	.071	24	.470	.064	27	.402	.099	51
+Planning	.339	.097	27	.456	.071	24	.394	.103	51
Total	.333	.085	51	.463	.067	51	.398	.101	102

(Planning F = .000, p = .989; Task type F = 72.78, p = .000; Interaction F = .822, p = .367)

Table 7. Lexical Word Range

	Descriptive			Narrative			Total		
	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N
-Planning	.467	.098	24	.685	.060	27	.582	.136	51
+Planning	.447	.113	27	.678	.075	24	.556	.151	51
Total	.456	.106	51	.682	.067	51	.569	.143	102

(Planning F = .593, p = .443; Task type F = 161.39, p = .000; Interaction F = .123, p = .727)

Table 8. Grammatical Word Range

	Descriptive			Narrative			Total		
	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N
-Planning	.284	.087	24	.359	.099	27	.324	.100	51
+Planning	.305	.098	27	.326	.074	24	.315	.087	51
Total	.295	.092	51	.344	.089	51	.320	.093	102

(Planning F = .112, p = .739; Task type F = 7.18, p = .009; Interaction F = 2.22, p = .139)

Table 9. Lexical-to-Grammatical Word Ratio

	Descriptive			Narrative			Total		
	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N
-Planning	.658	.113	24	.751	.083	27	.707	.108	51
+Planning	.699	.136	27	.824	.101	24	.758	.135	51
Total	.680	.126	51	.785	.098	51	.731	.124	102

(Planning F = 6.82, p = .010; Task type F = 24.71, p = .000; Interaction F = .540, p = .464)

Table 10. Lexical Density (%)

	Descriptive			Narrative			Total		
	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N
-Planning	37.56	5.40	24	41.65	2.82	27	39.73	4.67	51
+Planning	39.60	4.77	27	43.78	3.04	24	41.57	4.53	51
Total	38.64	5.13	51	42.65	3.09	51	40.65	4.67	102

(Planning F = 6.44, p = .013; Task type F = 25.30, p = .000; Interaction F = .003, p = .954)

The resulting planning effect yielded statistically significant results concerning lexical-to-grammatical ratio and lexical density – p=.010 (F=6.82) and p=.013 (F=6.44), respectively. However, no statistically significant results were found as measured in type-token ratio (F= .000, p=.989), lexical word range (F=.593, p=.443) and grammatical word range (F=.112, p=.739), respectively. In regard to the task type effect, statistically significant results were obtained on all the sub-measures of lexical richness – type-token ratio (F=72.78, p= .000), lexical word range (F=161.39, p= .000), grammatical word range (F=7.18, p= .009), lexical-to-grammatical ratio (F=24.71, p= .000), and lexical density (F=25.30, p= .000), respectively. There are no interaction effects in any of the measures reported above (p > .05). Therefore, hypotheses 7 and 8 received strong confirmation while hypotheses 4-6 were partly confirmed (i.e., only the task type effect appeared statistically significant).

LEXICAL ACCURACY

Lexical accuracy was measured by the total number of clauses with lexical errors. Table 11 below displays the results by planning condition and task type:

Table 11. Lexical Accuracy

	Descriptive			Narrative			Total		
	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N
-Planning	25.13	9.61	24	25.56	12.93	27	25.36	11.38	51
+Planning	12.94	10.81	27	19.88	8.80	24	16.21	10.42	51
Total	18.68	11.87	51	22.89	11.43	51	20.78	11.79	102

(Planning F = 17.60, p = .000; Task type F = 2.99, p = .087; Interaction F = 2.34, p = .129)

The results indicate that planning time results in a lower percentage of lexical choice errors, achieving a level of significance in which p = .01 (F = 17.60). The mean scores for planned tasks (regardless of type) are significantly lower than those for unplanned tasks (12.94 compared to 25.13 on descriptives, and 19.88 compared to 25.56, respectively). The total mean scores for no planning and with planning conditions are 25.36 and 16.21, respectively. There is, however, no significant effect for task type (F = 2.99, p = .05). There are no interaction effects (F = 2.34, p > .05). Thus, for the lexical accuracy measure, these results provide strong confirmation for one component of Hypothesis 9 concerned with planning, but no confirmation for the other, which is concerned with task type.

4 | DISCUSSION & CONCLUSION

The overall results indicated that planning time and task type are manipulable features of oral task performance. Through the regulation of these two constructs, a focus on lexical use may be induced, leading to oral production of varying lexical quality – lexical complexity (operationalized in syllabic range and lexical richness/variation), and lexical accuracy. More specifically, the results showed that the availability of planning time does not lead to greater lexical complexity in terms of syllabic range; however, it significantly increases lexical variation/richness in terms of lexical-to-grammatical ratio and lexical density (i.e., increased use of content/schematic vocabulary), and lexical accuracy as measured by lower number of clauses including lexical errors. As most previous related research used 'general' performance measures (Skehan, 2009b) rather than independent lexical measures, it is difficult to make fully accurate comparisons. For instance, the use of the general measure of accuracy includes lexical accuracy. Another source of difficulty is that lexical measures employed are in scarcity. In addition, Ellis (2009) expressed his concern by pointing out that "operational definitions have varied considerably" and that "[t]hese differences in the operational definitions are problematic as they make comparisons across studies difficult in some instances" (p. 475). Nevertheless, reference to previous research will be made where applicable to at least indicate the tendencies in which lexical use occurs. More direct comparisons will, however, be attempted regarding the previous few studies that employed similar types of lexical measures.

The overall results of previous research for complexity are mixed; however, there is a great amount of evidence that strategic planning results in more complex language (Ellis, 2009). Consistent with such evidence, which used

the three general measures of language production (i.e. complexity (i.e., grammatical complexity), accuracy and fluency), pre-task planning time leads to significant gains in terms of a certain set of lexical complexity measures, namely the two lexical variation/richness measures (i.e., lexical-to-grammatical ratio and lexical density), but not in terms of type-token ratio, which is a common measure of lexical complexity (Ellis, 2009), suggesting that planning time leads to a significantly higher level of content vocabulary use. It can be inferred that learners tend to use planning time to draw on more content words rather than grammar words as a result of schemata activation. This finding is in accord with the results of Bui (2019) and Gilabert (2007), which are among a limited number of planning studies that employed the lexical complexity measure. Bui (2019) reports that "pre-task planning raises lexical density" (p. 21). Similarly, the results of Gilabert's (2007) study indicated that planning led to significantly greater lexical richness, in contrast with earlier studies that found no effect for planning (e.g. Ortega, 1999; Yuan & Ellis, 2003). On the other hand, the availability of time to plan before task performance contributed positively to the degree of correctness of lexical use. From the perspective of the Trade-off Hypothesis (see Skehan, 2009b), an unprecedented trade-off was found within lexis. That is, in terms of lexical use, planning time leads to increased accuracy but lower complexity on several lexical measures, except for lexical variation and density measures. It should be noted here that lower lexical complexity is partially associated with planning time. It appears that greater lexical accuracy occurs at the cost of less varied lexical use as evident in the non-significant results in measures such as type-token ratio, lexical word range and grammatical word range. Interestingly, this trade-off within lexis seems to be in reverse direction to that discovered by Foster & Skehan (1996) and Skehan & Foster (1997) for the general measures of complexity (i.e., grammatical/syntactic complexity) and accuracy (i.e., grammatical/syntactic and lexical). They found that complexity and accuracy are in competition and that complexity is prioritized over accuracy when learners are afforded time to plan. The results can also be interpreted from the perspective of Cognitive Hypothesis (Robinson, 2001, 2003, 2005) to suggest that when L2 learners are given the chance to plan, they tend to focus on accurate as well as complex use of lexis, at least in terms of more varied and denser output. Clearly, more research is needed to verify first whether lexical complexity and accuracy can be mapped on to the general measures of complexity and accuracy, and next whether different sub-measures of lexical complexity and accuracy are equally sensitive and reliable in measuring lexical use, and lastly whether certain aspects of lexical use are more responsive to planning. As urged by previous literature, lexis needs to be a component of measures (Ellis, 2009; Skehan, 2009a, 2009b). Indeed, more studies that employ a variety of lexical measures (including particularly syllabic range used in this study) are needed before any independent component of lexis is comfortably incorporated into the existing repertoire of measures in task-based performance.

Task type, on the other hand, appeared to be more influential on the production of lexis. A similar crucial role of task type in L2 speaking has recently been suggested by Qui & Cheng (2021). However, the possible functions of task types have not been fully investigated (Qui, 2020). The two tasks (descriptives and narratives) were designed to elicit two distinct types of discourse – dialogic and monologic, respectively. Each type of discourse is characterized by a specific type of vocabulary put into use. The data showed that monologic discourse was associated with schematic vocabulary while dialogic discourse was connected to procedural vocabulary. Considering the measures in the present study, task type or discourse type led to significantly greater lexical complexity in terms of syllabic word range at all three levels (i.e., monosyllabic, disyllabic, polysyllabic), and significantly greater lexical variation/richness on all sub-measures (i.e., type-token ratio, lexical word range, grammatical word range, lexical-to-grammatical ratio, lexical density). For lexical accuracy, however, the effect of task type or discourse type did not reach a level of significance. The results are consistent with previous research which demonstrated a link between narrative tasks and higher complexity but lower accuracy and fluency (Skehan, 2009b). However, descriptive tasks, which can be seen as interactive tasks, did not advantage both accuracy and complexity (Skehan, 2009b), but only complexity.

A set of conclusions can be drawn based on the data. One major conclusion is that lexis is influenced more by task type than planning time. Particularly, discourse type as a consequence of task type has a strong effect on the type of vocabulary used. To illustrate, monologic discourse (triggered by narratives) leads to predominantly schematic vocabulary while dialogic discourse (triggered by descriptives) leads to predominantly procedural vocabulary. Schematic vocabulary is concerned with content words whereas procedural vocabulary involves grammar words. The implication here is that different types of vocabulary may be manipulated through the regulation of task type. In this way, a targeted aspect of vocabulary could be elicited, and therefore developed in

learners via task type regulation. This particular finding has great instrumental value for syllabus design as well as coursebook design where a set of tasks are mapped on to a specific type of vocabulary.

Another main conclusion is that lexical complexity, as measured in phonological complexity (word length in syllables), and lexical richness are largely determined by discourse type (stemming from task type) while lexical accuracy is determined to a great extent by planning. It is clear that the type of discourse L2 users engage in on a given task influences the complexity of lexis they put into use. Monologic narratives seem to generate more complex words in terms of number of syllables than the dialogic descriptives. Put another way, narrative tasks lead to lexically more complex (syllabic range) language than descriptive tasks. Likewise, the analysis of the data indicated that lexical richness was heavily influenced by discourse/task type. That is, narrative monologic discourse seems to lead to richer and more varied vocabulary use than the dialogic descriptives. Taking into account the results of lexical-to-grammatical ratio and lexical density, it can be concluded that lexical heaviness increases as a result of monologic discourse as well as planning. As for lexical accuracy, it is improved significantly more by planning than task type. In other words, L2 users tend to produce lexis with greater accuracy when they are afforded time for pre-planning. They would make better choices of words and use vocabulary more appropriately in the given context. This suggests that a focus on lexical complexity and lexical accuracy may be induced by manipulating discourse type and pre-task planning, respectively, and therefore that individual lexical measures may be selectively improved through task design. Obviously, this finding has crucial implications for Task-based Language Teaching (TBLT). In the light of such evidence, informed decisions can be made in selecting task features and conditions in designing TBLT syllabi in line with learning outcomes.

Overall, the general research question as to whether a focus on form can be induced through the regulation of task features (i.e., task type and planning time) was addressed adequately, revealing significant evidence of the relationship of task design and lexical use. The results demonstrated that learners' focus can be directed towards various types and aspects of vocabulary through systematic task design. Not only did the evidence confirmed the rightful call for lexis to be a component of general measures, but also supported the view that more research employing a variety of lexical measures is needed. Further research may specifically focus on the operationalization of between different lexical measures and how they interplay in oral language production. Evidence from such research will inform and consolidate choices made regarding lexical use within the framework of TBLT.

ACKNOWLEDGEMENTS

The current research article is based on one major component of the author's Ph.D. Thesis and its related data. A previous version of this article was presented as a conference paper at *14th International Conference of Applied Linguistics* on "Advances in Research on Language Acquisition and Teaching" in Thessaloniki, Greece.

STATEMENTS OF PUBLICATION ETHICS

The author declares that code of ethics for research and publication was closely adhered to. Ethical issues were given utmost importance and were handled with great care. Approval from ethics committee was not provided since the data reported in this research article were collected before 2020.

CONFLICT OF INTEREST

The author declares that there is no conflict of interest.

REFERENCES

Bui, G. (2019). Influence of learners' prior knowledge, L2 proficiency and pre-task planning on L2 lexical complexity. *International Review of Applied Linguistics in Language Teaching*, vol., 1-25. https://doi.org/10.1515/iral-2018-0244

Cambridge International Dictionary of English. (1995). Cambridge University Press.

Coles, M. (1982). Word perception, first language script and learners of English as a second language. MA dissertation, University of London, Birkbeck College.

Corson, D. (1995). *Using English words*. Kluwer.

Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition, 11*, 367-383.

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). What is lexical proficiency? Some answers from computational models of speech data. *TESOL Quarterly, 45*, 182-193.

DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 125-151). Cambridge University Press.

Derwing, T. M. (2017). L2 fluency development. In *The Routledge Handbook of Instructed Second Language Acquisition* (pp. 246-259). Routledge.

Duff, P. A. (1986). Another look at interlanguage talk: Taking task to task. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 147-181). Newbury House.

Ellis, R. (1987). Interlanguage variability in narrative discourse: style-shifting in the use of the past tense. *Studies in Second Language Acquisition, 9*, 1-20.

Ellis, R. (2005). Planning and task-based research: Theory and research. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 3-34). John Benjamins.

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics, 30*(4), 474-509.

Foster, P. (2020). Oral fluency in a second language: A research agenda for the next ten years. *Language Teaching, 53*(4), 446-461.

Foster, P. & Skehan, P. (1996). The influence of planning on performance in task based learning. *Studies in Second Language Acquisition, 18*(3), 299-324.

Foster, P. & Skehan, P. (1999). The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research, 3*(3), 215-247.

Foster, P. & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency and lexical diversity. *Language Learning, 59*(4), 866-869.

Gilabert, R. (2007). The simultaneous manipulation of task complexity along planning time and (+/- Here-and-Now): Effects on L2 oral production. In M. Garcia-Mayo (Ed.), *Investigating Tasks in Formal Language Learning* (pp. 44-68). Multilingual Matters.

Goldrick, M. (2014). Phonological processing: The retrieval and encoding of word form information in speech production. In M. Goldrick, V. S. Ferreira, & M. Miozzo (Eds.), *The Oxford Handbook of Language Production* (pp. 228-244). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199735471.013.015

Hamada, M. (2017). Influence of L1 orthography on multi-syllabic word recognition. *Journal of Psycholinguistic Research*, 46, 1101-1118.

Hobbs, J. (2005). Interactive lexical phrases in pair interview tasks. In C. Edwards et al. (eds.), *Teachers exploring tasks in English language teaching* (pp. 143-156). Palgrave Macmillan.

Johnson, M. D., Mercado, L. & Acevedo, A. (2012). The effect of planning sub-processes on L2 writing fluency, grammatical complexity, and lexical complexity. *Journal of Second Language Writing, 21*, 264-282.

Kahneman, D. (1973). *Attention and effort*. Prentice Hall.

Kang, S. & Lee, J.-H. (2019). Are two heads always better than one? The effect of collaborative planning on L2 writing in relation to task complexity. *Journal of Second Language Writing, 45*, 61-72.

Laufer, B. (1997). What's in a word that makes it hard or easy: some intralexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: description, acquisition, and pedagogy* (pp. 140-155). Cambridge University Press.

Laufer, B., & Nation, I. S. P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*, 307-322.

Levelt, W. J., Roelofs, A. & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(01), 1-38. https://doi.org/10.1017/S0140525X99001776

McLaughlin, B. (1990). Restructuring. *Applied Linguistics*, 11, 113-128.

Marslen-Wilson, W., Tyler, L., Waksler, R. & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review, 101*, 3-33.

Meara, P. (1990). A note on passive vocabulary. *Second Language Research, 6*, 150-154.

Meara, P. (1997). Towards a new approach to modelling vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 109-121). Cambridge University Press.

Meara, P., & Miralpeix, I. (2017). *Tools for researching vocabulary*. Multilingual Matters.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition, 20*, 83-108.

Melka Teichroew, F. J. (1982). Receptive vs. productive vocabulary: a survey. *Interlanguage Studies Bulletin* (Utrecht), *6*, 5-33.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nation, I. S. P. & Webb, S. (2011). *Researching and analyzing vocabulary*. Heinle Cengage Learning.

Nattinger, J. R. & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford University Press.

Ochs, E. (1979). Planned and unplanned discourse. *Discourse and Syntax, 12*, 51-80.

Ong, J. & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing, 19*, 218-233.

Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition, 21*, 109-148.

Pawley, A. & Syder, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. C. Richards and R. W. Schmidt (Eds.), *Language and Communication* (pp. 191-226). Longman.

Phillips, T. (1981). Difficulties in foreign language vocabulary learning and a study of some of the factors thought to be influential. MA dissertation, University of London, Birkbeck College.

Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communication tasks for second language instruction. In G. Crookes & S. M. Gass (Eds.), *Tasks and Language Learning: Integrating Theory and Practice* (pp. 9-34). Multilingual Matters.

Qui, X. (2020). Functions of oral monologic tasks: Effects of topic familiarity on L2 speaking performance. *Language Teaching Research, 24*(6), 745-764.

Qui, X. & Cheng, H. (2021). The effects of task types on L2 oral production and learner engagement. *International Review of Applied Linguistics in Language Teaching (IRAL)*, http://doi.org/10.1515/iral-2020-0128

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.

Revesz, A., Ekiert, M. & Torgersen, E. N. (2016). The effects of complexity, accuracy, and fluency on communicative adequacy in oral performance. *Applied Linguistics, 37*(6), 828-848.

Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for investigating task influences on SLA. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 287-318). Cambridge University Press.

Robinson, P. (2003). The cognition hypothesis, task design, and adult task-based language learning. *Second Language Studies, 21*, 45-105.

Robinson, P. (2005). Cognitive complexity and task sequencing: A review of studies in a Componential Framework for second language task design. *International Review of Applied Linguistics, 43*, 1-33.

Rodgers, T. (1969). On measuring vocabulary difficulty: an analysis of item variables in learning Russian-English vocabulary pairs. *IRAL, 7*, 327-343.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158.

Schmidt, R. W. (1992). Awareness and second language acquisition. *Annual Review of Applied Linguistics, 13*, 206-226.

Schmidt, R. W. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In Schmidt, R. (Ed.), *Attention and awareness in foreign language learning* (Technical Report 9), pp. 1-63. Second Language Teaching and Curriculum Center, University of Hawaii.

Schmidt, R. (2001). *Attention*. In P. Robinson (Ed.), Cognition and Second Language Instruction (pp. 3-32). Cambridge University Press.

Schmitt, N. (2010). *Researching vocabulary*. Palgrave Macmillan.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing, 18*, 55-88.

Segalowitz, N. (2003). Automaticity and second languages. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition*, pp. 382-408. Blackwell Publishing.

Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge University Press.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Skehan, P. (2009a). Lexical performance by native and non-native speakers on language-learning tasks. In B. Richards, H. M. Daller, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition: The interface between theory and application* (pp. 107-124). Palgrave Macmillan.

Skehan, P. (2009b). Modelling second language performance: Integrating complexity, accuracy, fluency and lexis. *Applied Linguistics, 30*, 510-532.

Skehan, P. & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1*(3), 185-211.

Stock, R. D. (1976). Some factors affecting the acquisition of foreign language lexicon in the classroom. PhD Thesis, University of Illinois.

Suzuki, M. (2017). Complexity, accuracy, and fluency measures in oral pre-task planning: A synthesis. *Second Language Studies, 36*(1), 1-52.

Tabari, M. A. (2016). The effects of planning time on complexity, accuracy, fluency, and lexical variety in L2 descriptive writing. *Asian-Pacific Journal of Second and Foreign Language Education, 1*, 10, 1-15. DOI: 10.1186/s40862-016-0015-6

Tabari, M. A. (2020). Differential effects of strategic planning and task structure on L2 writing outcomes. *Reading & Writing Quarterly*, *36*, 320-338.

Uchihara, T. & Clenton, J. (2020). Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research*, *24*(4), 540-556.

Ure, J. (1971). Lexical density and register differentiation. In G. E. Perren & J. L. M. Trim (Eds.), *Applications of Linguistics* (pp. 443-52). Cambridge University Press.

Webb, S. (2009). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC, 40*(3), 360-376. DOI: 10.1177/0033688209343854

Webb, S., Sasao, Y. & Ballance, O. (2017). The updated Vocabulary Levels Test. *ITL-International Journal of Applied Linguistics*, *168*, 33-69.

Wen, Z. E. & Ahmadian, M. J. (Eds.) (2019). *Researching L2 task performance and pedagogy: In honour of Peter Skehan*. John Benjamins. DOI: http://doi.org/10.1075/tblt.13

Widdowson, H. G. (1990). *Aspects of language teaching*. Oxford University Press.

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, *14*, 85-106.

Wood, D. (2001). In search of fluency: What is it and how can we teach it? *The Canadian Modern Language Review*, *57*(4), 573-589. DOI: 10.3138/cmlr.57.4.573

Yu, G. (2009). Lexical diversity in writing and speaking task performances. *Applied Linguistics*, *31*(2), 236-259.

Yuan, F. & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic production. *Applied Linguistics*, *24*(1), 1-27.