

**Task Complexity and Narrative Writing Performance:
The Effect of Proficiency**

Jelige: Nienna

ABSTRACT

The effect of task complexity on oral and written language production has been widely researched in recent years. Studies exploring the effect of task structure on L2 production are of special interest, as they have reported mixed results. The aim of the present study was to examine the possible interaction between task structure and proficiency level on L2 narrative writing performance. The participants were 46 high school students: 21 studying English as an L2 currently enrolled in an intensive language learning programme and 25 studying English as an L3. The participants' performance was examined on two narrative tasks differing in cognitive complexity. The findings indicate that the major difference between lower and higher proficiency learners lies in accuracy, lexical complexity/variety, and to some extent in syntactic complexity and fluency. The difference between the two groups in the structured task, however, was more substantial. With respect to the impact of task complexity on learners' written production, the results revealed that the cognitive demands of the tasks affect learners' of varying proficiency levels differently. The opportunity to plan the content of their narrative had a significant positive effect on the lexical sophistication and fluency and a minor beneficial influence on the accuracy of lower proficiency learners. The effect on higher proficiency learners, however, was less straightforward. While the less structured task elicited longer and to some extent syntactically more complex narratives, the structured task had a subtle positive effect on lexical variety and on some measures of syntactic complexity.

Table of Contents

Abstract.....	1
Table of Contents	2
Chapter 1: Introduction.....	4
Chapter 2: Review of Literature	6
2.1 Tasks in language pedagogy.....	6
2.1.1 Theories of task complexity and syllabus construction.....	7
2.1.2 Attention and performance: two contradictory views	8
2.1.2.1 Limited Capacity (Trade-off) Hypothesis	9
2.1.2.2 The Cognition Hypothesis	9
2.2 Written language production	10
2.3 Empirical research on cognitive task complexity.....	11
2.4 Research questions and hypotheses.....	13
Chapter 3: Method	14
3.1 Participants	14
3.2 Instruments	14
3.2.1 Narrative tasks	14
3.2.2 Proficiency test	15
3.3 Procedures	15
3.4 Analysis	16
3.4.1 Measures of task performance.....	16
3.4.1.1 Lexical complexity/diversity	16
3.4.1.2 Syntactic Complexity	16
3.4.1.3 Accuracy.....	17
3.4.1.4 Fluency	17
3.4.2 Statistical analysis	18
Chapter 4: Results and Discussion	19
4.1 Results	19
4.1.1 Proficiency test	19
4.1.2 Task performance on the two narrative tasks	19
4.1.3 Comparing the performance of the low-and high-proficiency groups	20
4.1.4 Comparing performance on the different task types	22

4.2 Discussion.....	24
4.2.1 Performance of low- and high-proficiency learners	24
4.2.2 The effect of the different task types	25
4.3 Conclusion.....	30
Chapter 5: Conclusions.....	31
5.1 Summary of main findings	31
5.2 Limitations of the study	33
5.3 Pedagogical implications and directions for future research.....	33
References	35
Appendix A – Cartoon strip task	39
Appendix B – Picture sequence task	40

Chapter 1: Introduction

The emergence of task-based language teaching produced a growing body of research investigating learners' performance on tasks from various perspectives. Numerous studies have sought to address the way task characteristics, most notably task complexity, direct learners' attentional resources to specific aspects of the target language during task completion. Understanding how learners' performance is affected by various task features has the potential to yield crucial insight into how tasks can be used to promote language learning. Furthermore, such research can serve as an empirical basis for syllabus design and teaching material development (Robinson, 2005).

A number of variables that conceivably affect learners' performance have been assessed; however, the findings about the way these influence output are not consistent. This is probably due to the vast number of different tasks and production measures used and the various proficiency levels and language backgrounds of the participants. Most of the studies examined intermediate (e.g., Robinson, 1995; Tavakoli & Foster, 2008) upper-intermediate (e.g., Kormos, 2011; Trebits, 2011) or advanced learners (e.g., Albert, 2008), but studies investigating learners with a mixed level of proficiency have been scarce. A notable exception to this is a series of studies undertaken by Kuiken and Vedder (2007a, 2007b, 2008) even though they acknowledged that the difference between the participants' L2 proficiency was not really substantial.

It is also important to note that even in those cases where the same tasks were used with participants at different proficiency levels (e.g., Albert, 2008; Kormos, 2011; Trebits, 2011) the results were markedly different. Manipulating task complexity along the [+/-structure] dimension, the findings of the aforementioned studies provide support for Skehan's (2009) claim that narrative tasks can place different cognitive demands on learners at distinct stages of language production. More specifically, a structured narrative with a predetermined content eases the conceptualisation of the output; however, due to the fact that narratives usually require the use of less frequent lexical items, it can make linguistic encoding difficult. A narrative with a loose structure, however, might exert a high cognitive demand on conceptualisation, but as Kormos (2011) and Trebits (2011) noted, by providing students with the opportunity to decide on the linguistic constructions and lexical items to be included in the narrative, it can ease linguistic encoding. Apart from this, however, the studies reported different results with respect to the effect of the

presence and absence of task structure on the different production measures, which suggests that there might be an interaction between proficiency level and the impact of these task characteristics.

The main purpose of this study, therefore, is to assess the relationship between L2 proficiency and task performance with the help of two tasks differing in cognitive complexity. The controlled study also makes it possible to compare the characteristics of the written narratives of lower and higher proficiency learners.

After providing an overview of the relevant literature on task-based research, the methodological aspects of the study will be described. Next, the results and discussion of the findings will be presented. The paper is concluded with a discussion of the limitations, pedagogical implications and future research possibilities.

Chapter 2: Review of Literature

2.1 Tasks in language pedagogy

A general dissatisfaction with traditional grammar-based syllabi, which aimed at improving declarative knowledge, led to the development and subsequent implementation of an alternative, communicative approach to language teaching (CLT) and syllabi design. Attention was drawn away from discrete grammatical structures and refocused on conveying meaning through language. In this framework, language production became not only the end, but also the means of second language development.

Long (1981, as cited in Bygate, 1999) suggested that during target language production, L2 learners are likely to encounter certain difficulties hindering smooth communication. In an attempt to overcome these communication breakdowns, L2 learners would modify their output in response to comprehension and confirmation checks. This process, known as “negotiation of meaning”, has been known to promote L2 learning ever since (see also Swain, 1995).

While teachers and SLA researchers generally agreed that communicative activities provide opportunities for spontaneous and creative language use, some also expressed their doubts about whether such communicative activities would be sufficient to drive interlanguage forward and foster language acquisition (Bygate, 1999). Based on Schmidt’s (1990) claims that “noticing” is necessary for language learning to occur, they were concerned that solely focusing on meaning and the effectiveness of communication will result in learners disregarding form completely. It was suggested that a heavy reliance on communication strategies can make communication effective to a certain extent, but that will not aid learners in making progress and might induce early fossilization (Skehan, 1996). Skehan and Foster (2001) drew attention to the lack of empirical research supporting the presumed role of negotiation in language development. They also noted that it is mainly gaps in learners’ lexical knowledge that trigger negotiation and not morphosyntactic features.

Consequently, the offspring of CLT, task-based language teaching (TBLT), while emphasizing the importance of meaning in interaction, also addresses the need for focusing on form; drawing learners’ attention to certain linguistic features through communicative language tasks. In TBLT the construct of “task” comes to the fore of syllabus construction (Skehan, 2003). A range of definitions have been proposed for the concept of “task”, each

emphasizing different features of the underlying construct: its connection to real-life activities (Long, 1985), its focus on conveying meaning (Nunan, 2004) and achieving an outcome (Candlin, 1987). Skehan (1996) provided the most comprehensive definition including all the characteristics highlighted by former wordings:

an activity in which meaning is primary, there is some sort of relationship to the real world, task completion has some priority, and the assessment of task performance is in terms of task outcome. (p.38)

2.1.1 Theories of task complexity and syllabus construction

Long (1985) noted that tasks, while differing in detail, share some characteristics in terms of their purpose or theme, which allows for their classification. He suggested that these general task-types help in the “subsequent principled selection of pedagogic tasks” (p. 92) and in the design of a final syllabus. According to Long, the basis of task sequencing, besides such general concerns as the time needed to solve the activities, should be the difficulty of the pedagogic tasks themselves. It should be noted, though, that Long did not provide a detailed description of those features of tasks that make them more or less difficult and implicitly influence learners’ performance.

The need to have a comprehensive taxonomy of task characteristics, first expressed by Long (1985), was addressed more specifically in subsequent years. In TBLT it is generally acknowledged that “pedagogic tasks should be developed and sequenced to increasingly approximate the demands of real-world target tasks” (Robinson, 2005, p.1). Real-life situations are more complicated than those encountered in a classroom setting; consequently, learners must learn how to cope with the high demands of complex tasks, as well. This idea led to the proliferation of research into those features of tasks that affect their complexity. The two most influential taxonomies have been developed by Skehan (1996, 1998) and Robinson (2001a, 2001b, 2005, 2007).

Skehan (1996, 1998) defined task complexity in terms of three distinct dimensions. The first of these, *code complexity*, concerns the linguistic demands of the task with regard to its grammatical and lexical difficulty. The second dimension, *cognitive complexity*, comprises two areas: *cognitive processing* and *cognitive familiarity*. The former refers to the amount of processing needed for task accomplishment and the amount, clarity and organisation of information that is provided in the input; while the latter pertains to familiarity with the topic, the discourse genre and the task. The final dimension, *communicative stress*, concerns anxiety-inducing factors such as time limit, modality,

number of participants, length of the texts used, the importance attached to the task and the amount of control learners have over the interaction. Skehan and Foster (2001) suggested that once the effect of these factors on language performance is determined by empirical research, such a scheme would be able to help in the grading and sequencing of pedagogic tasks.

Similarly to Skehan's model of task complexity, Robinson's Triadic Componential Framework (2001a, 2001b, 2005, 2007) also attempts to inform task sequencing decisions. Drawing on the findings of cognitive linguistics, differential and cognitive psychology and SLA, Robinson developed a theoretically grounded framework for characterising tasks (2005). He made a distinction between three independent aspects of tasks; *task complexity*, *task condition* and *task difficulty*; that all exert an influence on task performance.

Disregarding task condition, the interactive demands of tasks, and task difficulty, which stems from learners' cognitive abilities and affective dispositions, Robinson suggests that task complexity, the information processing demands of tasks, should be the sole basis for pedagogic task sequencing. This idea is based on the fact that while it is impossible to predict the impact of participant or affective variables in advance; that of task complexity can be reasonably foreseen. These cognitive demands are "fixed and invariant" features of tasks that make a task more or less demanding, and thus they can be a sound basis for preliminary task sequencing decisions (Robinson, 2001b, p. 30).

2.1.2 Attention and performance: two contradictory views

According to Swain's (1995) Output Hypothesis, active engagement in language production is crucial for L2 development, as output, while making language learners "notice" certain aspects of the L2, also gives them an opportunity to test and reflect on their hypotheses about the grammaticality and appropriateness of certain L2 structures as they "stretch their interlanguage to meet communicative needs" (p. 132). Attention to output, therefore, plays an important role in L2 learning. Consequently, the way task characteristics, especially task complexity, influence the allocation of learners' attention during task completion has been the principal focus of task-based research for years. Two contrasting hypotheses were put forward about the way attentional demands of tasks affect language performance, the Limited Capacity Hypothesis and the Cognition Hypothesis.

2.1.2.1 *Limited Capacity (Trade-off) Hypothesis*

Skehan (1996) identified the main goals of language teaching and learning as developing accuracy (language use conforming to L2 norms), complexity (sophisticated language use) and fluency (ability to communicate in real time), all of which require conscious attention on the learners' part. However, supposing that their attentional resources are limited, learners cannot pay attention to all three of these objectives simultaneously. According to Skehan, it is inevitable that one of the aforementioned goals will take priority over the others, which might affect performance in the other two areas negatively. Namely, Skehan and Foster (2001) proposed that cognitively demanding tasks will divert learners' attention away from linguistic forms, so that they can focus on the content of their message. This carries crucial implications for syllabus construction, as teachers should sequence tasks in a way that it allows learners to develop in all three areas of production.

Skehan (2009) also discussed the Trade-off Hypothesis in terms of Levelt's (1989) model of speech production. He argued that the structural complexity of the output is established at the Conceptualizer stage. A complex task, exerting a high cognitive demand on the first stage of speech production, would deplete learners' attentional resources and prevent them from paying close attention to message formulation. A less complex task, however, would demand less effort during conceptualization, making it possible to allocate more attention to the Formulator, the second component of the speaking process. Two performance dimensions, accuracy and fluency are linked to this stage; consequently, if more attention is available, they can be more effectively handled. Skehan suggests that, when resources are limited, the "natural tension" is between accuracy and complexity (p. 521). It should be noted, though, that in his model accuracy and fluency compete for the same attentional resources. Consequently, one would suggest a trade-off within the Formulation stage too; however, Skehan fails to elaborate on that (Trebits, 2011).

2.1.2.2 *The Cognition Hypothesis*

Task Complexity, as introduced in Robinson's Triadic Componential Framework (2001a, 2001b, 2005, 2007), can be manipulated along two distinct dimensions: *resource-directing* and *resource-depleting/dispersing* dimensions. Resource-directing task characteristics facilitate "noticing" those features of the L2 code that help learners in task completion. For instance, tasks involving causal reasoning require learners to use certain transition signals and conjunctive adverbs (e.g., therefore, consequently). Resource-

depleting variables, however, draw on students' attentional resources heavily, but do not make them focus their attention on specific L2 features. For example, reducing planning time places heavy burden on learners' working memory capacity.

Unlike Skehan, Robinson (2001a, 2005, 2007) suggested that there is no trade-off between accuracy and linguistic complexity. According to his Cognition Hypothesis, complex tasks can promote accuracy and complexity simultaneously, supposing that task complexity is increased along resource-directing variables. The high cognitive demands of tasks can make learners perform beyond the limit of their abilities. Increasing complexity along resource dispersing variables, however, results in poorer performance.

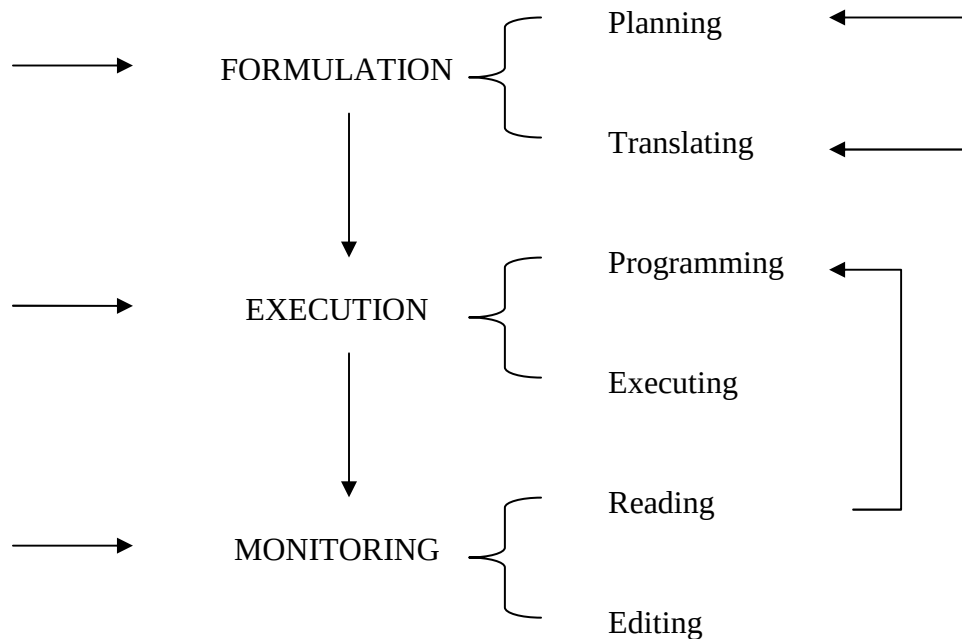
2.2 Written language production

Cognitive theories of writing, focusing on the mental processes underlying written language production, began to emerge at the end of the 1970s (for a recent review see Becker, 2006). All of these models addressed the complex nature of writing in an L1. For the time being, there is no generally accepted, comprehensive theory of L2 writing (Silva, 1993). However, since the stages of composing are essentially the same, L1 writing models are generally used to explain L2 writing processes.

The most influential theory of writing is probably that of Kellogg (1996). He identified three major stages of written language production: *formulation*, *execution* and *monitoring*. The first phase involves two processes: planning and translating. The former refers to goal setting, and generating and organising ideas into a coherent order, while the latter refers to the lexical and grammatical encoding of the conceptual representation of our message. The product of this encoding is stored in the writing buffer in a phonological form. During the programming phase of the second stage, execution, this phonological form is transformed into motor instructions, which are carried out during the executing phase, when the actual hand-written or typed text is created. The final stage, monitoring, involves re-reading the text, and after evaluation, revising it, if necessary. Monitoring, therefore, entails going back to the preceding stages of writing, making written language production a cyclical, recursive process.

Figure 1

The stages of writing



Note. From *Psycholinguistics: A resource book for students* (p. 114), by J. Field, 2003, New York, NY: Routledge.

The reason for choosing this model as the basis of this study is that one can easily link the stages of writing described by Kellogg to the stages of speech production defined by Levelt (1989). Since the effects of task characteristics on learner production have been discussed with respect to Levelt's model, having a writing model that allows for a direct comparison with that was essential.

According to Kellogg (1996), the three stages of formulation, execution and monitoring operate simultaneously, which imposes a heavy load on working memory. The human working memory capacity is limited, and L2 writers suffer greatly from this, as for them grammatical and lexical encoding is effortful and requires conscious attention. Since the subprocesses run in parallel, the conflicting demands they impose on working memory can interfere with each other and impair the quality of L2 writing (Trebits, 2011).

2.3 Empirical research on cognitive task complexity

In an attempt to test the Limited Capacity Model (Skehan, 1998; Skehan & Foster, 2001) and the Cognition Hypothesis (Robinson, 2001a, 2001b, 2005, 2007) and explain the effects of cognitive task complexity on learner performance, a large number of studies

were carried out in recent years. Most of these examined oral language production (e.g., Gilabert, 2007; Iwashita, McNamara, & Elder, 2001; Robinson, 1995; Skehan & Foster, 1999; Tavakoli, 2009; Tavakoli & Foster, 2008), but numerous studies examined written performance, as well (e.g., Ishikawa, 2007, Kuiken & Vedder, 2007a, 2007b, 2008). In this section, I would like to present empirical studies that investigated the effect of the degree of task structure on learners' output.

Albert's (2008) study explored the interaction between the complexity of narrative structure, creativity, aptitude and language proficiency on the oral narrative production of advanced learners of English. Statistical analysis revealed that the more complex task, where students had to invent the storyline, created more accurate narratives, while the less complex task, with a clear, inherent storyline, enhanced lexical diversity. Syntactic complexity remained invariant across the two tasks. With regard to individual differences, the findings indicate that those characterised by a higher level of proficiency and aptitude performed the less complex task with greater accuracy, the more complex with greater fluency, and both tasks with greater lexical variety. Interaction between the second individual variable examined, creativity, and task performance was moderate; but its association with the more complex task was stronger than with the less complex task.

Kormos (2011), adapting the tasks from Albert (2008), investigated the effect of task complexity on the written narratives of upper-intermediate foreign language learners and the quality differences between the narratives produced by them and native speakers of English. Her findings indicate that a tight structure urges students to use more abstract vocabulary and more temporal connectives to connect their sentences. The writings of foreign language learners differed from the natives' texts in lexical diversity, the use of infrequent words and clausal complexity.

Trebits (2011) examined the patterns of interaction between cognitive task complexity (+/- structure), oral and written task performance, two ability variables (i.e., aptitude and working memory capacity) and an affective variable, anxiety. The well-structured task elicited more varied vocabulary on the oral narratives, but failed to do so in the written narratives. In writing, the less-structured task resulted in greater syntactic complexity; however, it had a negative effect on accuracy and no effect on lexical diversity. Similarly to Skehan (2009), Trebits suggested that narratives can impose distinct demands on the different stages of language production. A loosely structured task, in which the story has to be created, places high demands on learners' during the planning phase. A structured task, however, in which there is no need to go beyond the mere

retelling of a story, requires less conceptualization but more effort in terms of formulating the message. These tasks are “input driven, and unforgiving in what needs to be covered” (Skehan, 2009, p. 517); consequently, they might push learners to use less frequent vocabulary items. According to the Cognition Hypothesis (Robinson, 2005), learner factors should play a more important role in performing cognitively more complex tasks. However, Trebits found that individual differences in working memory, aptitude and foreign language anxiety had a greater effect on the task characterised as less complex. This suggests that the structured task, although alleviating cognitive difficulties in terms of conceptualization, caused major complications with linguistic encoding.

It should be noted that the aforementioned studies (Albert, 2008; Kormos, 2011; Trebits, 2011) used the same tasks to elicit oral and written narratives and used similar production measures to analyse them, and still, they obtained different results. The fact that the participants’ proficiency level was different (advanced and upper-intermediate) suggests that L2 competence might have a crucial role in the way task characteristics affect performance.

2.4 Research questions and hypotheses

The study aimed to address the following research questions:

- 1) How does the written narrative performance of learners at various proficiency levels differ?
- 2) Does the influence of cognitive task complexity vary according to the level of L2 proficiency?

On the basis of previous studies the following hypotheses have been formulated:

Hypothesis 1: The supposedly less complex task with a tight structure will exert a higher cognitive demand on learners who are at a lower level of proficiency than the loosely structured task, and thus, it will have negative effects on all areas of production.

Hypothesis 2: High-proficiency learners, who are not likely to have difficulties with the vocabulary the tasks are designed to elicit, will benefit from the tight narrative structure and produce syntactically and lexically more complex output, while they will be impeded by the conceptualising demands of the loosely structured task.

Chapter 3: Method

3.1 Participants

The participants of the main study were 55 students studying English as a foreign language in three high schools in Budapest. They were between the ages of 16 and 18 studying in the second or third academic year of secondary grammar schools. The sample included 34 students studying English as a second foreign language and 21 studying English as a first foreign language.

The 34 students studying English as a second foreign language were recruited from three language groups of two different high schools. Even though their first foreign language was German, most of them started studying English before entering high school in language schools or with private tutors. Their years' of instruction ranged from six to nine. Consequently, their English language learning experiences were rather diverse. According to their teachers, the participants' proficiency level was B1 in the Common European Framework of Reference. This was supported in part by their performance on the proficiency test administered. However, the narratives of 9 students whose proficiency level was rated to be above B1 by the aforementioned test were excluded from the analysis.

The 21 students studying English as an L2 belonged to two language learning groups in the same high school, and they were enrolled in an intensive language learning programme. As part of this programme, in the so-called zero year, the students have nine 45-minute English lessons per week. In the upper years the amount of English lessons is reduced to five a week. The participants' proficiency level was intermediate; B2 in the Common European Framework of Reference, which was confirmed by the placement test. Thus, after eliminating some students' data in order to improve the homogeneity of the groups, the final sample consisted of 46 students, 20 males and 26 females.

3.2 Instruments

3.2.1 Narrative tasks

For the data collection, two tasks, both adapted from Albert's (2008) dissertation, were used. The cartoon strip task (see Appendix A) involved six pictures telling the story of a woman, whose broken-down car, instead of being repaired, was transformed into a carriage without horses by a wizard. The other, picture sequence task (see Appendix B)

consisted of six unrelated pictures: a book, a storm on sea, a boat, an island, the view of a street and an open door. In the case of the cartoon strip task, the Hungarian instruction asked the participants to write a narrative based on the chronological sequence of events depicted by the prompt. In the other, picture sequence task, students were asked to create a narrative using all of the pictures in a random order. In both tasks students could add extra information not shown in the pictures.

The two tasks were expected to place different cognitive demands on participants. In terms of Robinson's (2001a, 2001b, 2005, 2007) task classification, the second task is more complex, since it is less structured than the first and places higher cognitive demands on the students during the planning phase of the written production. The supposedly less complex task, however, might make greater demands on learners during the translating phase of the writing process, as for successful task completion, the use of certain low-frequency words might not be evaded (Skehan, 2009).

3.2.2 Proficiency test

The learners were administered the paper-and-pen version of the Oxford Quick Placement Test (UCLES, 2001). The test consists of 60 multiple choice items and should be completed in 30 minutes. It is made up of two parts: Part 1 (questions 1 – 40) has to be completed in 20 minutes, and Part 2 (questions 41 – 60) in 10 minutes. As the test proceeds, the questions get progressively more difficult.

3.3 Procedures

The participants completed the proficiency test and the two narrative tasks during their regular English classes in the presence of their teachers and the researcher. At first, the 30-minute long placement test was administered. Then, during another class, students were asked to write the two narratives without the help of dictionaries and textbook materials. They were required to write a minimum of 130 words for each task in 45 minutes. Since participants during the preliminary study managed to finish the two tasks in 30 minutes, a 45 minute lesson was considered to be more than enough to complete the narratives. In order to ensure that the sequence of the tasks does not have an effect on the results of the study, the order in which the narratives had to be completed was predetermined for each language group. One of the L2 groups and two L3 groups were instructed to start with the cartoon strip task, while the remaining L2 and L3 groups were asked to start with the picture sequence task. The instructions were given in the

participants' L1 in written format and were also read aloud by the researcher. A small sample of the narratives produced by the participants is included in Appendices C and D.

3.4 Analysis

3.4.1 Measures of task performance

In the present study, a number of production measures of four performance areas (accuracy, lexical diversity/complexity, syntactic complexity and length of the texts produced) were employed. In order to ensure reliability of the data analysis, a randomly selected sample of 20% of the narratives was coded by another rater. The inter-rater reliability, calculated using Pearson's correlation coefficients, was .99 for the syntactic complexity measures and .95 for the accuracy measures.

3.4.1.1 Lexical complexity/diversity

Lexical diversity was measured by Malvern and Richards' (1997) D-formula. Based on a mathematical probabilistic model, it was designed to overcome difficulties inherent in the type/token ratio, such as sensitivity to text length. The model is operationalised by the VOCED subprogramme of the CLAN software available from the CHILDES website (<http://childes.psy.cmu.edu/>). The D value is strictly concerned with lexical variety and does not indicate how sophisticated, difficult or rare words are used in the text. Accordingly, an additional measure of lexical complexity/diversity associated with lexical range was adopted. Nation's Range program (Heatley, Nation, & Coxhead, 2002) was used to identify what percentage of the words used in students' texts are part of the 1000 most frequent words, the second 1000 most frequent words, and what proportion is not part of the most frequent 2000 words but occur frequently in upper secondary school and university texts in a wide range of subjects. The latter is based on Coxhead's academic word list (2000); while the former two on West's "A General Service List of English Words" (1953).

3.4.1.2 Syntactic Complexity

Syntactic complexity, in line with Norris and Ortega's (2009) suggestions, was measured multidimensionally. Overall or general complexity was measured by the mean length of T-unit, a T-unit meaning an independent clause and all its dependent clauses, and by a sentence complexity ratio, which was calculated as the ratio of clauses to sentences. The reason for adopting the latter was that for lower proficiency learners using the

sentence as the unit of analysis is considered to be appropriate, as beginners often divide a single T-unit into more than one sentence (Ishikawa, 1995). Complexity via subordination, which is supposed to be a useful index of complexity at intermediate levels, was measured by the ratio of subordinate clauses to the total number of clauses (Norris & Ortega, 2009). Following Polio's (1997) guidelines, a clause was defined as a structure containing an overt subject and a finite verb. Since the operational definition of a clause does not treat participial phrases as clauses, a phrasal-level complexification measure, the mean length of a clause, was also adopted. A task-specific measure of syntactic complexity, the ratio of relative clauses, was also calculated. Trebits (2011) justified the use of this measure by arguing that relative clauses fulfil an important function in narratives, as they help in the introduction of characters, objects and places, and they are frequently used by L1 speakers when telling narratives.

3.4.1.3 Accuracy

In task-based research, accuracy is most often operationalised as the ratio of error-free clauses to the total number of clauses produced (Trebits, 2011). Consequently, this production measure was adopted in this research, as well. It should be noted, though, that this measure does not distinguish between clauses containing only one and those containing more than one error. Since with the lower proficiency learners it happened several times that the number of errors per clauses exceeded one, an additional accuracy measure, the average number of errors per 100 words, was also chosen.

In order to determine what counts as an error, Polio's (1997) detailed guidelines were followed. A clause, correct in isolation, was treated as erroneous provided that it did not fit into its preceding discourse. Spelling mistakes and errors in register were disregarded.

3.4.1.4 Fluency

Fluency is usually defined as "the capacity to mobilize one's linguistic resources in the service of real-time communication" (Skehan, 1996, p. 48). Calculating fluency in the case of written tasks, however, is problematic, and it is often disregarded (e.g., Kuiken & Vedder, 2007a, 2007b, 2008; Kormos, 2011; Trebits, 2011). Accordingly, there is no widely used fluency measure for writing; therefore, one of Ishikawa's (2007) length measures, the length of the text measured by the total number of words produced in English was adopted.

3.4.2 Statistical analysis

Data acquired from the analysis of students' written output were interpreted using SPSS 17.00 for Windows. After calculating descriptive statistics (mean and standard deviation), an independent samples t-test was employed to compare data collected from the lower-level and higher-level proficiency groups. Furthermore, a paired samples t-test was used to compare learners' performance on the two narrative tasks, treating the two groups of students separately.

Chapter 4: Results and Discussion

4.1 Results

4.1.1 Proficiency test

In order to make an estimate of the homogeneity of the lower and higher proficiency groups and to be able to exclude those whose command of the language is strikingly different from the rest of the group, the learners were administered the Oxford Quick Placement Test. The results of the test, with the scores ranging between 44 and 54, classified members of the higher proficiency group as intermediate and upper-intermediate learners (level B2 and C1 according to the Common European Framework of Reference). Accordingly, they constitute a rather homogeneous group. There was a greater variation between members of the L3 groups, with some students obtaining extremely high scores, 51 being the highest and 25 the lowest. Since it was essential to have two markedly different groups, those students from the L3 group whose proficiency scores were above 40 were excluded from the analysis. The proficiency test scores of those students whose written narratives were later analysed are presented in Table 2.

Table 1

Proficiency scores of the two groups of learners

	Number	Mean	SD
Lower-proficiency (L3) group	25	33.32	3.52
Higher-proficiency (L2) group	21	48.05	2.92

4.1.2 Task performance on the two narrative tasks

Table 3 displays the means and standard deviations of the measures of task performance for the lower and higher proficiency group respectively. With regard to lexical complexity/variety and fluency, there seems to be a lot of variation between students as standard deviations are relatively high. It should be noted, though, that this does not necessarily mean that there was a significant heterogeneity in the groups, as in similar studies (e.g., Albert, 2008; Kormos, 2011; Trebits, 2011) the standard deviations of these measures are equally high.

Table 2

Descriptive statistics of the production measures

Production measures	Lower proficiency group				Higher proficiency group			
	Cartoon description		Picture narration		Cartoon description		Picture narration	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
D value	57.35	14.24	57.08	17.67	68.35	16.23	61.17	14.87
Range wordlist 1	84.20	3.98	83.31	3.80	82.79	4.24	79.42	3.86
Range wordlist 2	7.86	2.26	8.67	2.72	7.72	2.97	11.54	3.08
Range wordlist 3	.51	.86	1.10	1.04	1.19	.94	1.3	1.03
Mean length of T-unit	8.77	1.92	9.20	1.41	9.70	1.33	9.85	1.68
Clauses per sentence	1.92	.42	1.99	.50	2.18	.52	1.93	.42
Subordination ratio	.24	.10	.26	.09	.29	.07	.24	.09
Mean length of clause	6.59	1.32	6.84	.92	6.86	.95	7.37	.92
Relative clause ratio	.04	.04	.06	.05	.04	.04	.08	.07
Ratio of error-free clauses	.59	.16	.63	.11	.82	.10	.79	.09
Average number of errors per 100 words	7.92	3.06	6.83	2.14	2.89	1.69	3.26	1.22
Text length	158.80	35.08	189.12	39.58	182.10	38.97	203.62	50.99

4.1.3 Comparing the performance of the low-and high-proficiency groups

The results of independent samples t-tests comparing the performance of the high- and low-proficiency groups are presented in Table 4 and 5. The fact that there were a number of statistically significant differences between them provides support for the validity of the grouping. Of the statistically significant differences detected, only the two accuracy measures were present in both tasks, indicating that the high-proficiency learners consistently produced more accurate output. The rest of the statistically significant differences between the two groups of learners, however, were on different production measures.

The number of significant differences was higher in the cartoon strip task and they could be observed with respect to lexical complexity/variety, fluency and syntactic complexity. The high-proficiency group used more varied vocabulary ($t(44)=2.448$, $p<0.05$), a higher percentage of rare words ($t(44)=2.543$, $p<0.05$) and produced longer narratives ($t(44)=2.133$, $p<0.05$). As for syntactic complexity, the high-proficiency group used more subordinate clauses ($t(44)=2.244$, $p<0.05$) and they tended to produce longer T-

units and incorporate a higher number of clauses into one sentence, although the last two of these observed differences failed to reach statistical significance.

Table 3

Independent samples t-test for performance measures on the cartoon strip task

	t	df	Sig (2-tailed)
D value	2.448	44	.018*
Range wordlist 1	-1.161	44	.252
Range wordlist 2	-.173	36.888	.863
Range wordlist 3	2.543	44	.015*
Mean length of T-unit	1.876	44	.067+
Clauses per sentence	1.904	44	.063+
Subordination ratio	2.244	44	.030*
Mean length of clause	.774	44	.443
Relative clause ratio	.288	44	.775
Ratio of error-free clauses	5.664	44	.001**
Average number of errors per 100 words	-7.042	38.513	.001**
Text length	2.133	44	.039*

+ $p < .08$. * $p < .05$. ** $p < .01$.

Interestingly, the number of statistically significant differences between the two groups' performance was reduced to four in the picture sequence task. The low-proficiency group, besides being less accurate, used more words from the most frequent 1000 English words ($t(44) = -3.433$, $p < 0.01$), while the high-proficiency group used more words from the second most frequent 1000 words ($t(44) = 3.353$, $p < 0.01$) and produced longer clauses, although the latter difference was not statistically significant. However, no variance with respect to lexical variety and fluency was found.

Table 4

Independent samples t test for performance measures on the picture sequence task

	t	df	Sig (2-tailed)
D value	.839	44	.406
Range wordlist 1	-3.433	44	.001**
Range wordlist 2	3.353	44	.002**
Range wordlist 3	.667	44	.508
Mean length of T-unit	1.440	44	.157
Clauses per sentence	-.494	44	.623
Subordination ratio	-.842	44	.404
Mean length of clause	1.931	44	.060+
Relative clause ratio	1.195	44	.239
Ratio of error-free clauses	5.620	44	.001**
Average number of errors per 100 words	-7.086	39.204	.001**
Text length	1.085	44	.284

+ $p < .08$. ** $p < .01$.**4.1.4 Comparing performance on the different task types**

Paired samples t-tests were conducted to compare the participants' performance on the two tasks, for the low- and high-proficiency learners separately. There were a number of statistically significant differences for both the low- and high-proficiency groups, although on different production measures. Therefore, the results, which are summarised in Table 6 and 7 below, indicate that the effect of cognitive task complexity is related to the participants' proficiency level.

In the case of the low-proficiency group, the loosely structured task increased the use of rare vocabulary items ($t(24) = -2.568$, $p < 0.05$) and relative clauses ($t(24) = -2.068$, $p = 0.05$), and elicited longer narratives ($t(24) = -3.385$, $p < 0.01$). Students also tended to be more accurate as measured by the average number of errors per 100 words. Although this trend is very clear, the difference did not reach statistical significance.

Table 5

Paired samples t-test comparing the low- proficiency group's performance on the two tasks

	t	df	Sig (2-tailed)
D value	.063	24	.950
Range wordlist 1	.962	24	.346
Range wordlist 2	-1.093	24	.285
Range wordlist 3	-2.568	24	.017*
Mean length of T-unit	-1.284	24	.211
Clauses per sentence	-.680	24	.503
Subordination ratio	-1.273	24	.215
Mean length of clause	-1.037	24	.310
Relative clause ratio	-2.068	24	.050*
Ratio of error-free clauses	-1.229	24	.231
Average number of errors per 100 words	1.992	24	.058+
Text length	-3.385	24	.002**

+ $p < .08$. * $p < .05$. ** $p < .01$.

For the high-proficiency group, significant differences are observed across the lexical complexity/variety measures; however, the differences point in different directions. Students used a smaller percentage of words from the 1000 most frequent English words ($t(20)=2.351$, $p<0.05$) and a larger percentage from the second most frequent 1000 words ($t(20)= -3.539$, $p<0.01$) in the loosely structured task. However, performance on the structured task produced more varied language even though the difference on this measure did not reach the level of significance. The presence of a clear task structure improved certain aspects of syntactic complexity, most notably the amount of subordination ($t(20)= 2.112$, $p<0.05$) and the sentence level complexity as indicated by the amount of clauses incorporated into a single sentence. On the other hand, the lack of task structure stimulated learners to produce longer clauses ($t(20)= -2.360$, $p<0.05$) and a greater amount of relative clauses ($t(20)= -2.622$, $p<0.05$). Similarly to the low-proficiency group, highly proficient learners also tended to produce longer narratives while performing the picture sequence task, although in their case the difference was not significant.

Table 6

Paired samples t-test comparing the high- proficiency group's performance on the two tasks

	t	df	Sig (2-tailed)
D value	1.898	20	.072+
Range wordlist 1	2.351	20	.029*
Range wordlist 2	-3.539	20	.002**
Range wordlist 3	-.388	20	.702
Mean length of T-unit	-.395	20	.697
Clauses per sentence	1.875	20	.075+
Subordination ratio	2.112	20	.047*
Mean length of clause	-2.360	20	.029*
Relative clause ratio	-2.622	20	.016*
Ratio of error-free clauses	1.656	20	.113
Average number of errors per 100 words	-1.159	20	.260
Text length	-2.001	20	.059+

+ $p < .08$. * $p < .05$. ** $p < .01$.

4.2 Discussion

4.2.1 Performance of low- and high-proficiency learners

More proficient learners are generally anticipated to produce more accurate, syntactically and lexically complex, and fluent output. Accordingly, it is not surprising to find these production measures differentiate between lower and higher proficiency learners. It is of interest, however, that some of these measures (e.g., D-value, subordination ratio, text length) failed to distinguish between the two groups in the second, supposedly more complex task.

It seems that the difference between the two groups was more pronounced in the cartoon strip task, whereas it levelled out to some extent in the picture sequence task. The low-proficiency group benefited from the lack of fixed narrative structure which resulted in a performance more similar to that of the high-proficiency group. It is probably due to the fact that the cartoon strip task had a prescribed storyline and accordingly a fixed set of vocabulary items that learners had to use in order to successfully complete the task. Consequently, they had limited opportunities for tailoring their message, forcing them to find a way to express their ideas to the best of their abilities. Narrating a story that

contained a fixed set of difficult vocabulary items required more cognitive effort from the low-proficiency group than it did from their more proficient peers.

These findings do not support Robinson's (2001a, 2001b, 2005, 2007) Cognition Hypothesis which states that performance on cognitively more complex tasks is more affected by individual differences, including L2 proficiency. Based on Robinson's Triadic Componential Framework, a task with inherent structure is cognitively less complex than a loosely structured task. However, in this study, English proficiency seemed to affect performance on the structured, thus cognitively less complex task to a greater extent.

4.2.2 The effect of the different task types

The main research question guiding the present study asked whether the influence of cognitive task complexity is different for low- and high-proficiency learners. Based on the results of previous studies it was hypothesized that the two types of tasks would impose a different amount of cognitive load upon the learners who are at different levels of proficiency. More specifically, it was hypothesized that the task with a tight structure would exert a higher cognitive demand on learners who are at a lower level of proficiency, deteriorating their performance with respect to accuracy, fluency, syntactic complexity and lexical variety. The high-proficiency learners, however, were expected to benefit from the predetermined storyline by producing syntactically and lexically more complex output and to be impeded by having to plan the content in the loosely structured task.

Paired samples t-tests comparing students' performance on the two tasks revealed that members of the low-proficiency group wrote longer narratives, used a higher number of relative clauses and a larger percentage of rare vocabulary items and tended to be more accurate on the picture narration task than on the cartoon strip task, thus providing support for Hypothesis 1. Higher proficiency learners, except for producing longer texts and more relative clauses in the loosely structured task, benefited from the two tasks in different ways than the lower proficiency group. The loosely structured task made learners incorporate fewer words from the most frequent 1000 English words and more from the second most frequent 1000 words into their narratives, but still, it made their output less varied lexically. The picture narration task led to an increase in clause length; nevertheless, the structured task also promoted higher syntactic complexity as measured by the ratio of subordinate clauses and the joining of clauses in a sentence. Accordingly, Hypothesis 2, which stated that higher proficiency learners would produce syntactically and lexically more complex output in the cartoon strip task, received only limited support.

The fact that both groups produced longer narratives in the picture sequence task is not surprising in the light of previous research, as Albert's (2008) study, which used the same tasks in oral mode, also demonstrated this. She explained this by postulating that completing the picture narration task, which did not provide learners with a fixed storyline, required a higher number of words to create a meaningful connection among the otherwise unrelated set of pictures. The cartoon strip task that depicted an event with a clear beginning, middle and end, however, made it possible for learners to be more concise in their writing.

The findings indicate a general tendency to use more relative clauses in the picture sequence task similarly to that of Trebits' (2011) study, which obtained the same results in the written mode. This is probably a result of the fact that relative clauses are used to identify or give information about the main characters, locations and events in a narrative. In the cartoon strip task, where all the main characters of the narrative were clearly depicted, students might not have felt the need to give a lot of background information about the protagonists and the location of the narrative. In the picture sequence task, however, participants were forced to be clearer about the referents in the story in order to provide readers with a sustained, meaningful set of events.

Similarly to Albert's (2008) advanced learners, the lower-proficiency learners had a tendency to be more accurate on the picture sequence task. She accounted for the greater accuracy in the more complex task by stating that the lack of clear storyline made it necessary for learners to pay attention to accuracy, as the imprecise use of tenses or personal pronouns would result in the story becoming disjointed and difficult to follow. In the cartoon strip task, however, in possession of the ordered set of pictures, the listener or reader could follow the narrative thread of the story even if it is rife with inaccuracies.

Since it was only the lower proficiency group that performed the tasks in a similar manner in the present research, other factors are likely to have influenced their accuracy. Skehan (2009) suggested that narratives in general encourage the use of less frequent lexical items, and due to the fact that the content of narrative tasks is rarely flexible, the use of these items probably cannot be evaded. It is possible that the lower proficiency learners have yet to master the use of certain less frequent words that were needed for successful task completion. Consequently, retrieval of those items or related items that could substitute the targeted words may have increased the processing demands of the task and consumed all the students' attentional resources, resulting in less accurate performance. The lack of tight narrative structure in the picture sequence task, however,

provided learners with more flexibility in completing the task. They could avoid the use of lexis that is beyond their competence and use expressions and structures that they have already mastered. The more fluent lexical retrieval gave them an advantage that resulted in more accurate performance.

Without the constraints of a tight narrative structure, low-proficiency learners were able not only to produce more accurate output but also to use a higher number of infrequent words. They could effectively exploit the freedom in planning the content of the narratives and use lexical elements that they were familiar with. In the studies that are most similar to the present research, Albert (2008), Kormos (2011) and Trebits (2011) all found that the presence of a structure facilitated the use of more sophisticated and varied vocabulary. It should be noted, though, that the participants' proficiency level in those studies was categorised as advanced and above-intermediate. Hence, it seems likely that the use of those less frequent lexical items (e.g., wizard, magic wand, carriage) that were difficult to avoid due to the content of the visual stimuli did not pose problems for them. The low-proficiency learners, on the other hand, benefited more from the loosely structured task, as it made it possible for them to avoid unknown words and tailor their message to incorporate those well-known words which are less frequent. The high-proficiency learners in the present study, whose proficiency level was probably somewhat below that of Kormos' and Trebits' sample, though used a smaller percentage of the most frequent 1000 words in English in the picture sequence task, tended to use a higher ratio of different words in the cartoon strip task. The presence of a tight story structure thus improved certain aspects of lexical complexity, while it deteriorated others. Still, they could benefit from the lower conceptualising demands of the cartoon strip task, and they could devote more attention to producing more varied output with less repetition.

Similarly to lexical complexity/variety, higher proficiency learners' performance on the two tasks produced mixed results concerning syntactic complexity. In accordance with Trebits's (2011) study, the picture sequence task resulted in longer clauses; however, in the present study the cartoon strip task also positively affected some aspects of syntactic complexity, namely the sentence complexity measures. The finding that the presence of a clear narrative structure elicited the use of more subordinate clauses might be explained by the fact that subordinate clauses in English are generally used to signal reason, result, purpose, cause, condition, manner, and to introduce indirect quotations. The visual stimuli in the cartoon strip task depicted a series of events with a clear chronological order and causality which might have encouraged the use of subordinating words to list events in

order of time and to give reason and result. Furthermore, the scene shown in the pictures also involved two protagonists presumably having a conversation with each other, which made learners use reported speech. This resource-directing characteristic of the cartoon strip task, however, did not push lower proficiency learners to use more subordination. It is possible that they found it hard to cope with the lexical demands of the task and did not have enough attention available to focus on syntactic complexity. As a result, they reported the characters' words or thoughts through direct speech rather than indirect speech.

In the case of the high-proficiency learners, the lack of structure in the picture sequence task had a positive influence on syntactic complexity as measured by clause length. An increase in clause length is usually thought to be caused by phrasal-level complexification; greater lexical density and the reduction of clauses into participial phrases (Norris & Ortega, 2009). According to Skehan (2009), such an increase in language complexity would be a response to the greater conceptualising demands of the task. The need to link a disjointed set of pictures in order to create a storyline imposes high cognitive demands on the planning stage of language production. This pressure on conceptualisation is expected to push learners to create a more complex pre-verbal message and subsequently a more complex outcome. It should be noted, though, that Skehan made these claims based on Levelt's (1989) model of speaking. This model assumes that the preverbal message is not yet linguistic, but solely the propositional content of the message. The syntactic frame of a sentence is only formed after grammatical encoding, when the retrieved lemmas trigger syntactic building processes (Kormos, 2006). Accordingly, a syntactically complex output would be the result of a pressure and push from the Formulator and not message planning alone, contradicting Skehan's claim that the Formulator is responsible only for accuracy and fluency and unrelated to syntactic complexity.

It would also be important to draw attention to the limitations of the clause length measure. It seems to be unreliable in the sense that longer clauses are not only the results of greater lexical density and nominalization, but grammatical subject deletion in coordinate clauses. As a result, longer clauses do not necessarily indicate a syntactically more complex outcome. At advanced levels of language proficiency, longer clauses might be inherently linked to phrasal elaboration, but in the case of intermediate learners it is not necessarily so; therefore, more accurate measures would be needed to assess this facet of complexification convincingly.

With respect to the two competing theories concerning the effect of task complexity on learner production, the findings of the present study support neither Skehan's (1996) Limited Capacity Model nor Robinson's (2001a, 2001b, 2005, 2007) Cognition Hypothesis. Based on Robinson's Triadic Componential Framework for task classification, the picture sequence task was expected to be more complex due to a resource-dispersing characteristic, the lack of tight narrative structure. The proposed effects of increasing task complexity along resource-dispersing variables were unanimously negative; both Skehan and Robinson suggested that it would result in the deterioration of accuracy, fluency and complexity. The lower-proficiency learners, however, seemed to benefit from the lack of structure which resulted in lexically more complex, more accurate and longer output. The higher-proficiency learners' accuracy was not affected by the task design, but the syntactic and lexical complexity and the fluency of their written compositions were; some positively by the cartoon strip task, others by the picture sequence task.

The findings, though, provide support for Skehan's (2009) claim, inasmuch as they suggest that one needs to separate the attentional demands on conceptualisation from the cognitive load imposed on linguistic encoding in the case of certain tasks. Nevertheless, his claims that the presence of a clear structure has a positive impact on fluency and accuracy but not on syntactic complexity; and that the need to organize, manipulate and integrate presented information leads to greater syntactic complexity at the expense of accuracy and fluency were not substantiated. It should be noted, though, that when formulating this idea, Skehan did not take L2 proficiency into consideration. The results of the present study seem to suggest that there is a complex interaction between the command of the language and the different cognitive demands of tasks. For lower-proficiency learners linguistic encoding is more effortful; therefore, any demanding lexical pressures can be seriously detrimental to them. They can, however, turn to their advantage and benefit from task complexity brought about by the lack of tight task structure, as it makes it possible to relieve the pressure on the translating stage of language production. On the other hand, pressures on planning and translating the message probably place similar attentional demands on learners who are at a higher level of proficiency, which could explain that in some aspects of language they performed better on the picture sequence and in others on the cartoon strip task.

The results also seem to justify Skehan and Foster's (2001) claim that task complexity is related not only to the cognitive complexity of the tasks but also to "code

complexity”, the linguistic demands of tasks. They acknowledge that the language focus of tasks also affects learners’ performance. As opposed to this, Robinson (2001a) claims that task content is “independent of the cognitive demands of tasks, and can be adjusted to learner factors such as proficiency level without affecting the design and sequencing of tasks themselves” (p. 300) and accordingly should not play a role in sequencing decisions in syllabus design. The findings of the present study, however, suggest that the linguistic demands of tasks contribute to task difficulty and, therefore, should not be disregarded.

4.3 Conclusion

The aim of this chapter was to compare the narrative output of the lower and higher proficiency groups on the two tasks and to see how their performance varied as a result of the task design. The results of independent samples t-tests indicate that production measures tapping into all the different areas of performance (accuracy, syntactic complexity, lexical complexity/variety, and fluency) were able to distinguish low-proficiency learners from their higher proficiency peers; however, the difference between the two groups was more substantial in the cartoon strip task than in the picture sequence task. As far as task design is concerned, participants in the present study seemed to be affected by the presence and absence of structure in the narrative tasks. For the low-proficiency learners, a positive influence of the lack of structure was captured by four measures: the use of a higher number of infrequent words, the increased number of relative clauses, the longer texts and fewer errors produced in the picture sequence task. In the case of the higher proficiency learners, the presence of a tight narrative structure facilitated the use of a more varied vocabulary and a greater number of subordinate clauses; while the less structured task elicited longer narratives and clauses and a greater number of relative clauses and made learners use fewer words from the 1000 most frequent English words.

The findings do not lend support to Skehan’s (1996) Limited Capacity Model, nor to Robinson’s (2001a, 2005, 2007) Cognition Hypothesis. However, the idea that task design can exert a variable impact on different stages of language production was sustained. The results also show that participants’ proficiency level interacted with the task design and that both factors affected their performance. Hence, when categorising tasks according to cognitive complexity, the proficiency level of the target language learners has to be taken into account, as well.

Chapter 5: Conclusions

5.1 Summary of main findings

The main objective of the study was to investigate the interaction of task complexity and L2 proficiency on the participants' L2 narrative writing performance. Task complexity was manipulated along a resource-dispersing variable of Robinson's (2001a, 2001b, 2005, 2007) Triadic Componential Framework, the degree of task structure. Given a set of 6 sequential pictures and 6 unrelated pictures as visual prompts, a group of B1 and B2 level learners were asked to write two narratives. Two research questions were addressed about the characteristics of the sample and the joint effects of task complexity and L2 proficiency.

The first research question examined the way lower- and higher-proficiency learners differ in terms of written narrative performance. The results revealed a number of statistically significant differences between the lower-intermediate (B1 level) and the intermediate (B2 level) students on both tasks; however, the difference between the two groups was more marked on the structured cartoon strip task than on the picture sequence task. The higher-proficiency students consistently produced more accurate output. The production measures related to syntactic complexity, lexical complexity/variety and fluency were also able to distinguish between the two groups, although not in a uniform manner. More varied vocabulary, a higher percentage of rare words and subordinate clauses, and longer T-units and narratives characterised the written output of the higher-proficiency learners on the cartoon strip task. In the case of the less structured task, however, no significant differences were observed across the lexical variety and fluency measures. Some of the lexical complexity measures (use of the most frequent 1000 and second most frequent 1000 English words) and a syntactic complexity measure (clause length) could differentiate between the two groups.

The second research question considered whether there is a difference in the effect of task complexity on the performance of lower and higher proficiency learners. The task with a tight narrative structure, due to its lexical constraints, was expected to impose a higher cognitive demand on learners who are at a lower level of proficiency. On the other hand, the higher proficiency learners, whose linguistic resources are less limited, were expected to find having to plan the content in the loosely structured task cognitively more demanding. The hypotheses were partly confirmed, as B1 level learners profited from the

lack of task structure inasmuch as it made them use a higher number of infrequent words and relative clauses, and produce longer texts and fewer errors. For the B2 level learners, however, both the presence and absence of task structure improved certain aspects of written language production. The task with a predetermined storyline encouraged the use of a more varied vocabulary and a higher number of subordinate clauses. The less structured task, however, produced longer narratives and clauses, a greater number of relative clauses and it made learners use fewer words from the 1000 most frequent and more from the second 1000 most frequent English words.

The two major theories of task complexity, Skehan's (1996) Limited Capacity Model and Robinson's (2001a, 2001b, 2005, 2007) Cognition Hypothesis, suggest that the less structured task, being more complex along resource-dispersing variables, would result in less accurate, complex and fluent performance. The findings do not lend support to these theories, as the performance of the low-proficiency group was pushed remarkably, and that of the high-proficiency group to some extent by the cognitively more complex task. However, Skehan's (2009) idea that the amount of cognitive load imposed on learners can be different at different stages of language production was clearly demonstrated. Tight structure makes conceptualisation easier, as learners do not have to invent the content of the narrative. However, the lack of flexibility in completing the task can make linguistic encoding more difficult, especially for learners who are at a lower level of proficiency. A loosely structured task, on the other hand, is difficult in terms of planning, as students are required to construct a meaningful plot out of the unconnected pictures. In exchange, linguistic encoding is considerably eased as learners are free to use structures that they have already mastered while avoiding those which are beyond their current level of competence.

Skehan's (2009) claim, however, that having to invent the story would result in more complex output, while accuracy and fluency would be positively affected by the tight narrative structure was not supported. The results of the present research suggest that when discussing the effect of the differing cognitive demands of tasks one needs to take learners proficiency level into account, as well. For lower proficiency learners the cognitive demands on linguistic encoding can be much more severe than the demands placed on conceptualisation. In the case of higher proficiency learners, for whom linguistic encoding is less effortful, however, the two types of cognitive demands they encounter in these tasks seem to be equally challenging.

5.2 Limitations of the study

A number of limitations inherent in the present research need to be acknowledged and considered when interpreting the results. First of all, the extent to which the findings can be generalised is restricted by the characteristics of the sample. The participants, both at B1 and B2 level, were studying in prominent secondary schools in the capital. Accordingly, they form a specific subsample of the Hungarian high school population. It is important to keep in mind that the conclusions of the study can only be valid for the population from which a sample is drawn.

Secondly, a major limitation of the study is that the number of participants was relatively small. As the data analysis is labour intensive and time-consuming, task-based research is usually conducted with similar sample sizes. Still, the findings indicate a number of clear tendencies and near significant differences, which might have reached the level of statistical significance on a larger sample.

Finally, a more sophisticated language proficiency test tapping into different areas of performance might have been useful in characterising the sample's proficiency level more accurately. Furthermore, it would have made it possible to examine the correlations between different aspects of task performance and English proficiency measures.

5.3 Pedagogical implications and directions for future research

The main goal of the present research was to contribute to the different task complexity frameworks which help in grading and sequencing tasks. According to Robinson (2001a), cognitive task complexity should be the sole basis of sequencing decisions in syllabus design. The present findings, however, suggest that not only task complexity, but also language proficiency should be taken into account when implementing tasks in the classroom. It seems that different tasks impose a different amount of cognitive load upon learners who are at different levels of proficiency. Low proficiency learners seemed to be able to perform to the best of their abilities in the less structured task, which made them produce more fluent, complex and accurate output. However, the more demanding cartoon strip task could also prove useful in learning and practicing the use of specific vocabulary items. The two tasks differing in cognitive complexity made higher proficiency learners focus on different aspect of performance. Accordingly, combining these task types would direct learners' attention to all performance areas and lead to a balanced development.

The present research necessarily raised some questions that further research will need to address. Besides conducting a similar study with a larger sample size in order to increase validity of the findings, further studies could explore the impact of other types of written and oral tasks on lower and higher proficiency learners' performance. An issue that was not addressed in this study is whether certain ID variables (e.g., aptitude, motivation, anxiety) contribute to task completion and how they interact with language proficiency. It would also be interesting to explore participants' perceptions of the influence of task structure through interview or questionnaire studies.

REFERENCES

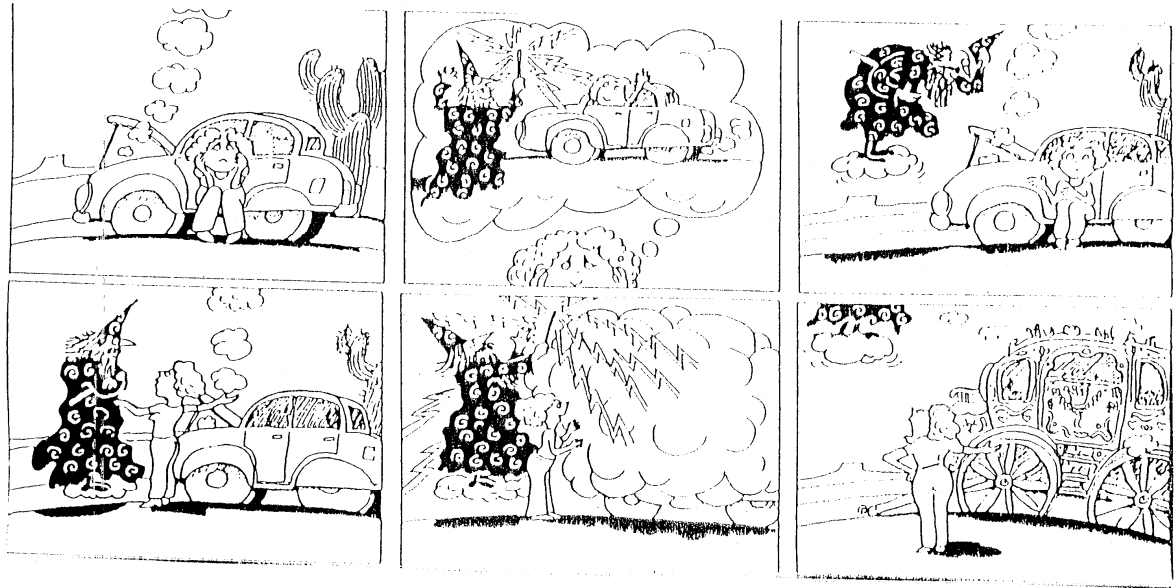
- Albert, Á. (2007). *Creativity and oral narrative task performance: A study of first year English majors* (Unpublished doctoral dissertation). Eötvös Loránd University, Budapest, Hungary.
- Becker, A. (2006). A review of writing model research based on cognitive processes. In A. Horning & A. Becker (Eds.), *Revision: History, theory, and practice* (pp. 25-49). West Lafayette, IN: Parlor Press.
- Bygate, M. (1999). Task as context for the framing, reframing and unframing of language. *System*, 27(1), 33-48.
- Candlin, C. (1987). Towards task based language teaching. In C. Candlin & D. Murphy (Eds.), *Language learning tasks*. Englewood Cliffs, NJ: Prentice Hall.
- Carroll, J. B. (1993). *Human cognitive abilities*. New York, NY: Cambridge University Press.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Field, J. (2003). *Psycholinguistics: A resource book for students*. New York, NY: Routledge.
- Gilabert, R. (2007). The simultaneous manipulation of task complexity along planning time and [+/- Here-and-Now]: Effects on L2 oral production. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 44-68). Clevedon, UK: Multilingual Matters.
- Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). Range [Computer Software]. Available from <http://www.victoria.ac.nz/lals/resources/range>
- Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing*, 4(1), 51-69. doi:10.1016/1060-3743(95)90023-3
- Ishikawa, T. (2007). The effect of manipulating task complexity along the [+/- Here-and-Now] dimension on L2 written narrative discourse. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 136-156). Clevedon, UK: Multilingual Matters.
- Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information processing approach to task design. *Language Learning*, 51(3), 401-436. doi:10.1111/0023-8333.00160

- Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 57-71). Mahwah, NJ: Lawrence Erlbaum.
- Kormos, J. (2006). An overview of theories of first language production. In *Speech production and second language acquisition* (pp. 3-11). Mahwah, NJ: Lawrence Erlbaum.
- Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, 20(2), 148-161. doi:10.1016/j.jslw.2011.02.001
- Kuiken, F., & Vedder, I. (2007a). Cognitive task complexity and linguistic performance in French L2 writing. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 117-135). Clevedon, UK: Multilingual Matters.
- Kuiken, F., & Vedder, I. (2007b). Task complexity and measures of linguistic performance in L2 writing. *International Review of Applied Linguistics in Language Teaching*, 45(3), 261-184. doi:10.1515/IRAL.2007.012
- Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, 17, 48-60. doi:10.1016/j.jslw.2007.08.003
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Long, M. (1985). A role for instruction in second language acquisition: Task-based language teaching. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp.77-99). Clevedon, UK: Multilingual Matters.
- Malvern, D. D., & Richards, B. J. (1997). A new measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving models of language* (pp. 58-71). Clevedon, UK: Multilingual Matters.
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555-578. doi:10.1093/applin/amp044
- Nunan, D. (2004). *Task-based language teaching*. New York, NY: Cambridge University Press.
- Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, 47(1), 101-143. doi:10.1111/0023-8333.31997003

- Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning*, 45(1), 99-140.
- Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287-318). Cambridge, UK: Cambridge University Press.
- Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22(1), 27-57. doi:10.1093/applin/22.1.27
- Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching*, 43(1), 1-32. doi:10.1515/iral.2005.43.1.1
- Robinson, P. (2007). Criteria for grading and sequencing pedagogic tasks. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7-27). Clevedon, UK: Multilingual Matters.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 127-158.
- Silva, T. (1993). Toward and understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27(4), 657-677. doi:10.2307/3587400
- Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, 17(1), 38-62. doi:10.1093/applin/17.1.38
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford, UK: Oxford University Press.
- Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36, 1-14. doi:10.1017/S026144480200188X
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510-532. doi:10.1093/applin/amp047
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49(1), 93-120. doi:10.1111/1467-9922.00071

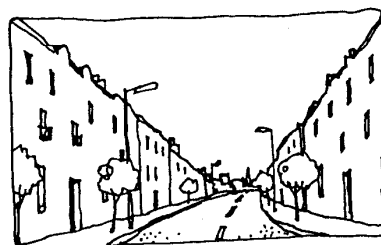
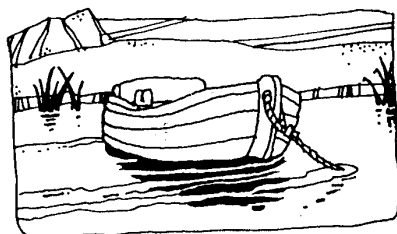
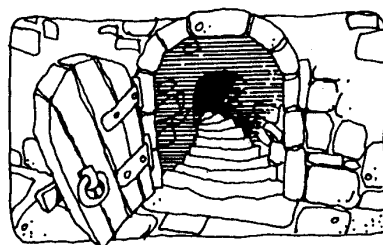
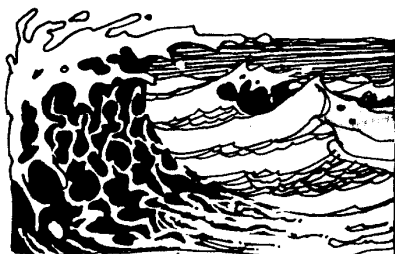
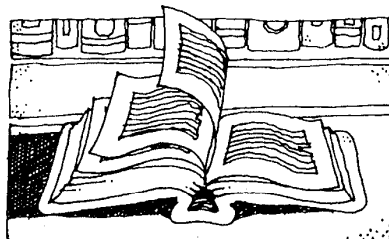
- Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183-205). Cambridge, UK: Cambridge University Press.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125-144). Oxford, UK: Oxford University Press.
- Tavakoli, P. (2009). Assessing L2 task performance: Understanding effects of task design. *System*, 37(3), 482-295. doi:10.1016/j.system.2009.02.013
- Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 58(2), 439-473. doi:10.1111/j.1467-9922.2008.00446.x
- Trebits, A. (2011). *Task complexity and second language narrative production: Effects of individual differences in working memory, foreign language aptitude and anxiety* (Unpublished doctoral dissertation). Eötvös Loránd University, Budapest, Hungary.
- UCLES (2001). *Quick placement test: Paper and pen pack*. Oxford, UK: Oxford University Press.
- West, M. (1953). *A general service list of English words*. London, UK: Longman.

Appendix A – Cartoon strip task



Please write a story of at least 130 words which tells the reader what happened in this cartoon. You may also add stages not shown by the pictures.

Appendix B – Picture sequence task



You will find six pictures in the envelope. Please write a story of at least 130 words, which includes all the elements depicted by the pictures. You must use all the pictures from the envelope, but you may also add extra information if you wish.